(2)

AD-A260 816

Effect of Aptitude on the Performance of Army Communications Operators

John D. Winkler, Judith C. Fernandez,
J. Michael Polich

DTIC
ELECTE
FEB 23 1993
S E D

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

93-02874

93 2 12 183

RAND | ARROYO CENTER

The research described in this report was sponsored by the United States Army, Contract No. MDA903-91-C-0006.

Library of Congress Cataloging in Publication Data

Winkler, John D., 1953–
Effect of aptitude on the performance of army communications operators / John D. Winkler, Judith C. Fernandez, J. Michael Polich.
p. cm
"Prepared for the United States Army."
"R-4143-A."
Includes bibliographical references.
ISBN 0-8330-1212-6
1. United States. Army—Occupational specialties. 2. United States. Army—Electronic technicians—Ability testing. 3. United States. Army—Electronic technicians—Rating of. 4. United States. Army—Communication systems. I. Fernandez, Judith C., 1948– . II. Polich, J. Michael. III. Title.
UB337.W56 1992
623'.73—dc20 91-42075
CIP

RAND is a nonprofit institution that seeks to improve public policy through research and analysis. Publications of RAND do not necessarily reflect the opinions or policies of the sponsors of RAND research.

Published 1992 by RAND
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138

R-4143-A

Effect of Aptitude on the Performance of Army Communications Operators

John D. Winkler, Judith C. Fernandez,
J. Michael Polich

Prepared for the United States Army

Accesion For		
NTIS CRA&I	X	
DTIC TAB	☐	
Unannounced	☐	
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

DTIC QUALITY INSPECTED 3

RAND

Approved for public release; distribution unlimited

PREFACE

This report documents results of RAND Arroyo Center research on the linkage between the aptitude of enlisted personnel and their ability to operate and troubleshoot tactical communications systems. The purpose of the research was to improve the ability of the Army to set appropriate performance standards and to estimate the effects of personnel quality levels on Army operational performance. This study was one of several research efforts on soldier performance conducted by RAND and the U.S. Army Research Institute. The results should be of interest to manpower analysts in the Army, the Office of the Secretary of Defense, and the other services, as well as to policy analysts interested in the relationship between the aptitude of military enlisted personnel and their performance on combat-related tasks.

THE ARROYO CENTER

The Arroyo Center is the U.S. Army's federally funded research and development center (FFRDC) for studies and analysis operated by RAND. The Arroyo Center provides the Army with objective, independent analytic research on major policy and management concerns, emphasizing mid- and long-term problems. Its research is carried out in four programs: Strategy and Doctrine, Force Development and Technology, Military Logistics, and Manpower and Training.

Army Regulation 5-21 contains basic policy for the conduct of the Arroyo Center. The Army provides continuing guidance and oversight through the Arroyo Center Policy Committee (ACPC), which is co-chaired by the Vice Chief of Staff and by the Assistant Secretary for Research, Development, and Acquisition. Arroyo Center work is performed under contract MDA903-91-C-0006.

The Arroyo Center is housed in RAND's Army Research Division. RAND is a private, nonprofit institution that conducts analytic research on a wide range of public policy matters affecting the nation's security and welfare.

Lynn E. Davis is Vice President for the Army Research Division and Director of the Arroyo Center. Those interested in further information about the Arroyo Center should contact her office directly:

Lynn E. Davis
RAND
1700 Main Street
P.O. Box 2138
Santa Monica CA 90407-2138

SUMMARY

BACKGROUND AND OBJECTIVE

The modern Army's battlefield operations depend to a great extent on the rapid availability of communications. During combat, communication among dispersed units is made possible by communications operators in mobile Army signal centers, which typically represent "nodes" in a network handling multichannel signals. This report describes RAND research to assess the performance of signal operators and to link that performance to personnel aptitude, as measured by scores on the Armed Forces Qualification Test (AFQT), the Defense Department's test of general aptitude.[1]

This study was one of several research efforts sponsored by the U.S. Army to develop quantitative analyses based on objective measurement of soldier and unit performance. The primary purposes of these research efforts were to improve the Army's ability to set appropriate performance standards and to develop quantitative estimates of the link between personnel aptitude and Army operational performance.

This research examined duty tasks performed by military occupational specialty (MOS) 31M, Multichannel Communications Equipment Operator, whose members operate communications systems providing division- and corps-level command and control. A key feature of this research was its examination of wartime-related tasks using the facilities of a high-fidelity tactical communications simulator to provide an objective and systematic test of performance. The research examined two principal functions:

- *Communications system operation*: the ability to establish an operating communications network, a task that requires interaction and teamwork among individuals at different communications facilities.

[1]AFQT scores, along with scores on other composite scales of occupational aptitude, are derived from the Armed Services Vocational Aptitude Battery (ASVAB), the written test used to screen military applicants. The AFQT measures general mental aptitude as percentile scores from 1–99 normed on the U.S. youth population. AFQT categories are in turn defined by AFQT scores as follows: category I, percentiles 93–99; category II, 65–92; category IIIA, 50–64; category IIIB, 31–49; category IV, 10–30. Category V persons, percentiles 1–9, are excluded by law from military service.

- *Communications system troubleshooting*: the ability to isolate faults in multichannel communications systems, also a task requiring teamwork.

For both functions, we examined graduates of Advanced Individual Training (AIT) at the Army Signal Center, Fort Gordon, Georgia. For communications system operation, we also report results for soldiers from active-duty Signal Corps units. The report covers two major sets of analyses. The first set examined the performance of Signal Corps operators in establishing and operating a functional communications system. We examined the effect of different levels of AFQT scores on the likelihood that a three-person group will successfully operate the system. We also examined the performance of individual operators on related tasks (those involved in preparing equipment for operation), showing the extent to which proficiency depends on individual AFQT score. The second set of analyses examined the performance of groups of operators in isolating faults in malfunctioning communications systems. Taken together, these results imply that AFQT score has a direct, consistent effect on the ability of communications personnel to provide effective battlefield communications to Army units.

COMMUNICATIONS SYSTEM OPERATION

Group Performance

We used the Reactive Electronic Equipment Simulator (REES)—a high-fidelity, computer-controlled simulation facility—to obtain measures of a group's ability to install and operate a realistically configured communications network, such as that connecting a division command post to division artillery. Groups of three operators (two at terminals and one in a relay position) were assigned to communications nodes in a controlled experiment to provide a realistic wartime mix of aptitude levels and to allow us to analyze performance for varying levels of group aptitude and experience. We examined the performance of 240 such three-person groups, which were formed from a set of 720 new graduates of the Signal Center's AIT course for MOS 31M, Multichannel Communications Equipment Operator. We also examined 84 three-person groups (252 31Ms) from active-duty signal battalions. The tests were performed between September 1988 and April 1989.

We wanted to assess how group outcome (i.e., success or failure in establishing a functioning system) was affected by various soldier characteristics, such as aptitude, experience, and demographic and educational background. To make this assessment, we used regres-

sion models that allowed us to predict the effects of varying levels of aggregate group aptitude (measured as the average AFQT score of the group's members) while controlling for other differences.

The results of our analyses demonstrate that the average AFQT score of a three-person group of operators is an important determinant of group success in system operation when measures of demographic background, education, and military experience are controlled. The groups with lower average AFQT scores were significantly less likely than those with higher average AFQT scores to establish a functioning communications system involving two terminals, one relay, and two 12-channel systems. Some of our results fcr members of active-duty signal battalions can be used to illustrate these differences.

The model predicts that for randomly selected groups of three soldiers in which the average AFQT score is at the midpoint of category IIIA, 63 percent will successfully operate the system within the allotted time. However, if group aptitude is reduced so that the average AFQT score falls to the midpoint of category IIIB (and all other factors are held constant), the prediction is that only about 47 percent of the groups will be successful. The model thus suggests that for groups of soldiers performing the specified task, the effect of lowering the average AFQT score from the midpoint of category IIIA to the midpoint of category IIIB is to reduce the probability of successful system operation by 16 percentage points.[2]

The results of our analyses further indicate that the aptitudes of all group members contribute to the probability that the group will operate the system successfully. For example, we found out how the probability of successful system operation relates to the number of group members (out of three) whose scores fall within categories I through IIIA.[3] Our models suggest that as the number of group members in categories I through IIIA increases, the group is more likely to operate the system successfully. Each additional "high-scoring" member improves the probability that the group will succeed by about 8 percentage points.

---

[2]For groups of three AIT graduates, the model predicts that the probability of successful system operation for groups whose average AFQT score falls at the midpoint of category IIIA will be 15 percentage points higher than that for groups whose average AFQT score falls at the midpoint of category IIIB.

[3]Individuals with scores in this range are commonly referred to as *high-aptitude* personnel.

Individual Operator Performance

For measures of the individual proficiency of terminal operators, we drew on data collected in earlier RAND studies that permitted us to examine the relationship between AFQT score and individual proficiency. The initial objective of these earlier studies was to examine the effectiveness of alternative training strategies at the Signal Center, but they also provided the basis for subsequent analyses to examine the effects of AFQT score on performance.

Our analyses showed that for a variety of tasks and equipment, personnel with higher AFQT scores are significantly more likely than their lower-scoring counterparts to install their assemblages correctly. The first of these analyses involved 340 AIT students in MOS 31M who were tested in the REES facility. The results show that higher-scoring operators are more likely to accomplish the initial steps of the system operation task (preset and cabling of the AN/TRC-145 terminal). Other analyses examined 336 trainees in MOS 31Q (Tactical Satellite/Microwave Systems Operator), whose performance was measured in hands-on tests administered by objective assessors who were unaware of the soldiers' AFQT scores. Again, the results show that the higher the AFQT score, the more likely the 31Q operator will be to successfully prepare the equipment for operation (i.e., perform alignments and adjustments of tropospheric scatter radios) according to Army technical standards. Across these various tasks, the models predict performance differences of approximately 5 to 8 percentage points by AFQT category. Thus, for instance, if individual AFQT scores fall from the midpoint of category IIIA (close to current levels) to the midpoint of category IIIB (through a reduction in accession standards, for example), we would expect the ability of operators to successfully perform the three preset, alignment, and adjustment tasks to decrease by 5 to 8 percentage points per task.

COMMUNICATIONS SYSTEM TROUBLESHOOTING

The final set of analyses examined the ability of groups to perform a second major function: isolate faults that interfere with the proper operation of a communications system. For this test, we used the REES facility to introduce malfunctions in an operating system composed of two terminals and two relays. The faults were selected so that teamwork would be a significant factor in identifying the sources of the problems and determining the appropriate corrective actions. As in our test of system operation, groups of three AIT graduates were formed and assigned to nodes to provide a range of group aptitude. Each group of AIT graduates received six malfunctions to

isolate.[4] Altogether, the test involved 187 three-person groups, representing 561 individuals.

The analyses used regression models to predict group success at isolating faults given various levels of group aptitude and controlling for differences in the experience and the demographic and educational background of the group members.[5] The results show that as the average AFQT score of the group increases, so does group success at isolating faults to the correct assemblage and component. For example, the model predicts that 60 percent of randomly composed groups of operators whose average AFQT score falls at the midpoint of category IIIA will find two or more bugs (the median value). If the average group AFQT score falls to the midpoint of category IIIB, however, the same outcome is expected only 43 percent of the time. Thus, the decrease in performance predicted for a decline in average AFQT score from the midpoint of category IIIA to the midpoint of category IIIB is 17 percentage points.

CONCLUSIONS

Our results provide considerable evidence that AFQT score has a direct effect on the ability of signal operators to provide usable battlefield communications to the Army. The effects of AFQT score on performance are remarkable in several respects. They manifest themselves across a range of tasks and situations. We found the effects of AFQT score to be statistically significant for group performance on two different tasks—system operation and system troubleshooting—and for individual proficiency on tasks involving very different kinds of equipment and functions for two different operator MOSs (31M and 31Q). Moreover, for system operation, the AFQT score effects were found to be similar for both AIT graduates and unit members, even though the latter group has more experience.

Perhaps more important, we found evidence that AFQT score has a sizable effect on group performance. In general, we observed that

---

[4]Each group received three troubleshooting trials of 10 minutes each with two faults inserted per trial. The assignment of faults was counterbalanced: "bugs" appeared in two different nodes per trial; over the course of the test, each examinee received an equal number of bugs. The faults consisted of two malfunctions inserted in a radio transmitter, three in a radio receiver, and one in a multiplexer. The symptoms of the faults ranged from red alarm lights to incorrect meter readings to audible cues (alarms or buzzers) that failed to sound. In all three trials, a fault could be diagnosed most quickly if team members cooperated with one another.

[5]Specifically, we used ordered polytomous logistic regression to predict the probability of finding some minimum number of bugs, e.g., one or more, two or more.

groups that are on average "smarter" outperform other groups. The analyses suggest that the effects are additive: each member contributes to successful performance on communications tasks that require interaction and coordination among group members.

Together, these results indicate that a change in accession standards that causes a shift in average AFQT category from IIIA (close to current levels) to IIIB will substantially reduce the probability of operator success in operating and troubleshooting communications systems. The results are significant—reductions of 16 to 17 percentage points in the probability of successful operation and troubleshooting—and imply that a reduction in average AFQT levels would carry a penalty in battlefield performance and readiness, an effect that should be considered in making budget and resource allocation decisions.

ACKNOWLEDGMENTS

The authors owe a debt of gratitude to General Maxwell Thurman (retired), who initiated the soldier performance research effort during his tenure as Commander of the U.S. Army Training and Doctrine Command (TRADOC) and who offered key support, assistance, and encouragement throughout the effort. We further acknowledge the assistance and support provided by Major General Theodore Stroup, who sponsored this research as Deputy Chief of Staff for Resource Management, TRADOC, and as Director of Military Personnel Management, Office of the Deputy Chief of Staff for Personnel. We also received helpful advice from Mr. Thomas Edwards, Assistant Deputy Chief of Staff for Training (Plans) of TRADOC, along with able assistance from our project officers during this study: Lieutenant Colonels David Block, Bob Donoho, and Ken Martell.

The research reported here would not have been possible without the support of the U.S. Army Signal School and Center, Fort Gordon, Georgia, with active interest from its former Commanding General, Major General Leo M. Childs, and former Deputy Commanding Generals, Major General Alfred J. Mallette and Brigadier General Robert E. Gray. We offer our special thanks to Mr. Glenn Swan, Mr. Irv Ingraham, Sergeant First Class Bobby Waters, and Master Sergeant Ken Rhodes of the Proponency Office, and to Mr. Dewey Plunkett, Director of the Transmissions Operations Department.

We also acknowledge important contributions made by our RAND colleagues. We are especially grateful to Mr. James Steward, who oversaw the tests of system operation and troubleshooting described in this report. Capable research assistance was provided by Kirsten Alesch, Marilyn Yokota, Jim Vernon, and Tom Martin, while invaluable statistical advice was provided by Neal Thomas and Robert Bell. Jan Hartman and Bob Young provided valuable programming assistance, and Melissa Huff, Velda DeCosentine, and Linda Tanner prepared the manuscript and briefing slides. RAND colleagues Beth Asch and Rob MacCoun provided helpful technical reviews of an earlier draft. The authors, of course, are solely responsible for any shortcomings.

CONTENTS

FIGURES

TABLES

1. INTRODUCTION

BACKGROUND AND OBJECTIVE

This report presents the results of research linking the aptitude of soldiers to their ability to perform tasks important to success in combat. The issue at the heart of the research is the relationship between enlistment standards and job performance. Historically, the Department of Defense (DoD) and the military services have sought to recruit the most talented individuals possible. A key criterion for enlistment is "general aptitude," as measured by the Armed Forces Qualification Test (AFQT), which is part of the Armed Services Vocational Aptitude Battery (ASVAB), the qualifying examination for admission to the military.

An above-average AFQT score and a high school diploma are the characteristics of the "high-quality" recruit most desired by the military.[1] But the appropriate quality levels for recruiting objectives and the recruiting resources needed to meet those objectives have long been issues of policy debate. The military naturally seeks to recruit the most talented people by drawing primarily from the top half of the aptitude distribution (categories I through IIIA), and Congress has established legislative minimum standards for "lower-quality" recruits. Currently, no more than 20 percent of enlistees may have scores in AFQT category IV.[2] Congress and the services have often disagreed, however, on what is an acceptable "mix" of higher- and lower-quality enlistees, particularly as to what is an acceptable balance between the percentage of enlistees falling in category IIIB and the percentage falling in categories I through IIIA.

A key reason for interest in this issue is the cost of recruiting a higher-quality force. Upon graduating from high school, the higher-aptitude individuals sought by the services generally have attractive postsecondary-education and employment options available to them, so more resources are required to attract these individuals to military

[1]AFQT scores are represented as percentile scores normed on the U.S. youth population. They are used to define five AFQT categories: I, percentiles 93–99; II, 65–92; III, 31–64; IV, 10–30; V, 1–9. Category III is further subdivided into categories IIIA (50–64) and IIIB (31–49). The percentages of youths qualifying in each category (1980 reference population) are: I, 8 percent; II, 28 percent; III, 34 percent; IV, 21 percent; V, 9 percent (Wigdor and Green, 1989).

[2]Applicants with AFQT scores in category V (1–9) are excluded by law from military service.

service. The problem has been especially acute for the Army, which recruits the largest number of new personnel each year and is often still perceived as having predominantly low-skill jobs, despite the infusion of considerable high-technology equipment in recent years.

In fact, defense budgets have incorporated special Army programs to attract high-quality recruits since the beginning of the All-Volunteer Force in 1973. During most of this period, the Army has maintained substantial incentives for high-quality personnel, including special educational benefits and cash bonuses, and has aimed its advertising at high school graduates and college-bound youth. Such programs are widely credited with helping to achieve a generally high level of quality across DoD throughout the 1980s.

However, the need for these costly recruiting resources has persistently been questioned, given the scarcity of evidence linking high-quality personnel with improved combat performance and readiness.[3] It seems likely that in the coming climate of defense-budget constraints, policymakers will focus increasing attention on the services' stated requirements for high-quality personnel and their associated costs. Potential reductions in the size of U.S. military forces could intensify this concern. Some will argue that in a smaller Army, maintaining a flow of high-quality accessions should be easier, and recruiting incentives may be less important given the current high level of high-quality personnel already in the force. Counter-arguments will surely be made. For example, some will assert that a larger proportion of the personnel in a smaller Army should be high quality in order to maintain a cadre for training a surge of new personnel in a crisis or to keep a very ready, quick-response contingency force capable of multiple missions. In such an Army, soldiers could be asked to perform a wider variety of tasks, and there will be less opportunity to practice those tasks because of constraints on budgets and exercises.

These broad issues about the future were beyond the scope of the current study, but they illustrate the perennial importance of questions about personnel quality. Underlying all such questions is the uncertainty about the value of high-quality personnel: How much do they contribute to an armed force's combat capability and readiness? To

[3]For example, in its report accompanying the FY 1988 military authorization bill, the House Appropriations Committee directed DoD to develop new methods of linking the educational background and aptitude of recruits to the ability of units to perform their operational missions (Department of Defense, 1987). In response to this requirement, the U.S. Army Training and Doctrine Command (TRADOC) sponsored the program of research that includes the study reported on here.

address this uncertainty so that the Army's scarce resources for recruiting and other purposes can be effectively used, analyses are needed to establish the relationship between Army personnel quality and performance.

Ideally, such analyses should link broad categories of resources (such as personnel quality and training opportunity) to objective, quantified assessments of individual and unit performance on wartime-related tasks. Also in the ideal, such analyses should deal with tasks germane to unit combat capability across a range of Army functions or operating systems. Thus far, however, the research has fallen short of these ideals. Most analyses of personnel quality requirements have been based on the minimum aptitude levels that individual recruits need to pass initial skill training courses.[4] Recently, the Army's long-range job performance measurement project, Project A, has sought to provide more definite connections between various recruit characteristics and certain performance measures on specific critical tasks.[5] However, this effort has focused more closely on individual task performance than on broader outcomes relevant to combat performance. The research has generally failed to link individual aptitude to unit outcomes or to examine the unit performance of units directly.[6]

This report presents the results of RAND research intended to develop improved data bases and analyses for such broad-based performance assessment. Our objective was to produce empirically based, quantitative estimates of the relationships between soldier aptitude and the job performance of crews and small units. The research described here was performed in the area of Army communications, one of several different functional areas addressed by the overall research effort.[7]

---

[4]TRADOC schools set minimum aptitude entry standards for their individual training courses, and training standards are established in various TRADOC-published soldier's manuals, ARTEP-MTPs (Army Training and Evaluation Programs and Mission Training Plans), and related publications.

[5]See, for example, Campbell (1990).

[6]An exception is research linking the AFQT scores of tank crews to gunnery accuracy on test ranges (Scribner et al., 1986), which demonstrates significant improvements in accuracy as the AFQT scores of the tank commanders and gunners increase. Such an effect, however, has not yet been demonstrated for other military functions and tasks.

[7]A related RAND study concerns the performance of air defense units (Orvis, Childress, and Polich, 1991). Studies in other areas have been carried out by the U.S. Army Research Institute (Graham, 1990a,b; Horne, Reilly, and Schopper, 1990) and by TRADOC schools and centers (e.g., Schopper, Johnson, and Burley, 1990; Silva, 1990).

Army communications was selected as an area for research because of its importance in battlefield command and control. During wartime, the Signal Corps provides the infrastructure that permits communication between individuals and units on the battlefield. A large number of enlisted personnel work, usually in teams, in the military occupational specialties (MOSs) that operate and maintain communications equipment. The equipment with which they work—transmitters, receivers, switching equipment, and other electronic gear—is among the more complicated equipment in the Army's inventory. The aptitude of personnel working with this equipment would be expected to affect individual performance and the performance of communications teams, which in turn would affect the ability of the units to communicate during combat.

However, the relationship between AFQT score and the performance of communications personnel had been examined in only a few previous empirical studies. Two studies conducted at the Signal Center examined noncommissioned officers (NCOs). The first of these (Donaldson, 1985) employed a written test to measure job knowledge of members of MOS 31C30, Signal Channel Radio Operator Supervisor. The test was administered to 41 active-duty signal personnel, and the results showed a moderate positive correlation between AFQT score and job knowledge ($r = 0.40$). The second study examined students taking the end-of-course test in the basic NCO courses for six signal MOSs (LeGree and Sanders, 1990). It also found a moderate correlation between AFQT score and test score. Finally, a moderate correlation between AFQT score and the hands-on task performance of individual Single Channel Radio Operators (MOS 31C10) was found in analyses performed as part of Project A (Office of the Assistant Secretary of Defense, 1990).

These earlier studies demonstrate that AFQT score and the performance of enlisted communications personnel are positively related. However, they do not examine performance differences between individuals in AFQT categories of policy interest—i.e., the performance expected from "high-aptitude" personnel (categories I through IIIA) versus that expected from "low-aptitude" personnel (categories IIIB and IV)—while accounting for other differences, such as experience and education. The practical significance of differences in aptitude for performance in combat cannot be inferred easily from these studies. None of them have examined how performance is affected as individuals who differ in aptitude interact to achieve a common combat objective. The current research was undertaken in part to obtain such additional insights, as well as to provide improved

estimates of the relationship between broad categories of personnel and training resources and combat-related outcomes.

RESEARCH APPROACH

This research focused on the primary tasks performed by MOS 31M, Multichannel Communications Equipment Operator, whose members operate communications systems providing division- and corps-level command and control. A key feature of the analysis was its assessment of *group* performance at two wartime-related tasks:

- *Communications system operation:* the ability to establish a functioning multichannel communications network, a task that requires interaction and teamwork among individuals at different communications facilities.
- *Communications system troubleshooting:* the ability to isolate faults in multichannel communications systems, also a task requiring teamwork.

For both functions, we examined graduates of Advanced Individual Training (AIT) at the Army Signal Center, Fort Gordon, Georgia. For the test of communications system operation, we also examined the performance of personnel serving in several active-duty Army signal units.

Operator performance was tested using the Reactive Electronic Equipment Simulator (REES), a high-fidelity communications simulator located at the Army Signal Center. This simulator has four signal nodes, each representing a signal center such as is found at a division or corps headquarters. The REES's simulation capabilities provide a realistic environment for systematic and unbiased evaluation of operator performance. Its computer system provides a mechanism for recording student data for subsequent analysis.

We also examined earlier tests of *individual* performance that offered us the opportunity to assess possible AFQT score effects. During a series of previous RAND studies on alternative training approaches for signal personnel, systematic data were collected on soldier performance for two MOSs, 31M and 31Q (Tactical Satellite/Microwave Systems Operator, the members of which operate satellite communications equipment and other high-technology gear). Put together, the data from these earlier studies cover a range of important tasks, some measured in the REES and some measured using hands-on tests with actual equipment. We reanalyzed these data to specify the differ-

ences in performance levels that can be expected among individuals with varying AFQT scores.

REPORT OUTLINE

The remainder of this report presents our research methodology and results for the tests of operation and troubleshooting of communications systems. Section 2 describes the overall approach taken in the research. Section 3 presents the results of analyses linking the effects of personnel quality to the operation of tactical communications systems. We describe how the average aptitude of groups of three operators influences group success in operating a communications system. Then, in Sec. 4, we use data collected in earlier research on Army signal students to examine the effects of personnel quality on individual proficiency in performing presets, alignments, and adjustments of multichannel communications equipment. Section 5 presents analyses linking the average aptitude of groups of three operators to their performance in isolating faults in a communications system. Finally, Sec. 6 describes the implications of our findings for Army policy on accession standards and mission readiness. Supporting analytical tables are provided in Apps. A and B.

2. RESEARCH APPROACH

This section describes the primary research approach we used to assess the relationship between aptitude and proficiency for enlisted personnel who operate multichannel communications equipment, a key activity in the Army Signal Corps. Most of the studies discussed used the facilities of the Army Signal Center's Reactive Electronic Equipment Simulator (REES). Two broad types of tasks were assessed:

- Install and operate multichannel communications equipment as necessary for a division or corps to communicate.
- Isolate faults and identify corrective steps in troubleshooting communications systems.

The research design allowed us to derive statistical estimates of how personnel quality affects the performance of these tasks, which in turn implies the availability of communications facilities to commanders at division and corps levels. Below, we describe the occupational specialty examined, the characteristics of the simulator from which most of the performance measures were drawn, and the procedures used to select individuals and form groups for testing.[1,2]

OCCUPATIONAL SPECIALTY

The Army Signal Corps provides the communications backbone connecting units on the battlefield. As described in Army doctrine (Department of the Army, 1977a) each division has an associated signal battalion that provides internal communications within the division and links the division to its subordinate units.[3] Division units, and division artillery or maneuver battalions, each have an assigned communications element that moves with them, as do division headquarters elements. Within communications units are teams that establish and maintain communications links, providing commanders

[1]Methodological details specific to the tests of communications system operation and troubleshooting are described in Secs. 3 and 5, respectively.

[2]At the request of the Signal Center, we also helped to design a hands-on test of crew proficiency in installing antennas. This test, however, did not have the systematic evaluation properties of the REES simulator test, and its results are not reported here.

[3]A signal brigade provides the analogous function for the corps.

with the ability to command, control, and coordinate all types of subordinate units.

Among the various signal-related military occupational specialties (MOSs), career management field (CMF) 31 provides the personnel who operate the tactical communications equipment used in units. One of four CMFs in Army communications, CMF 31 provided 49,000, or 75 percent, of the nearly 65,000 enlisted spaces authorized in the Signal Corps in FY 1989, when this study took place. Thirteen entry-level MOSs and four advanced supervisory MOSs were included within CMF 31 (U.S. Total Army Personnel Command, 1989).

This study examined the performance of operators in a key MOS, 31M, on tasks essential to the operators' wartime mission. As described in Army doctrine (Department of the Army, 1978), 31M personnel hold important positions at command posts and area signal nodes within headquarters and area signal battalions. One of their jobs is to link command posts at corps/divisions with adjacent corps/division command posts, subordinate headquarters, and other important elements, such as division artillery and air defense batteries.

The members of MOS 31M install, operate, and perform preventive maintenance checks and services and unit-level maintenance on multichannel communications equipment and such related equipment as antennas and generators (Department of the Army, 1989). The initial skill level (31M10) emphasizes installation, operation, and maintenance of equipment. Higher levels (e.g., 31M20 and 31M30) have the added responsibility of supervising other team members.

The equipment used by MOS 31M is arrayed as individual *components* (e.g., radio transmitters, receivers, communications security devices), *assemblages* of integrated components, and *shelters* consisting of assemblages within their assigned vehicles. Assemblages are attached to portable generators located outside the shelters and to a movable antenna. The major categories of assemblages used by 31Ms include the so-called high-capacity, medium-capacity, and low-capacity equipment.[4] Among the most frequently used is the low-capacity assemblage, AN/TRC-145, which is found in the REES. The AN/TRC-145 (pictured in Fig. 2.1) is a radio terminal set containing

[4]The "capacity" of equipment denotes the number of different individual channels, or bands of frequencies, on which communication can be established. Higher-capacity equipment is usually concentrated at the highest echelons of command.

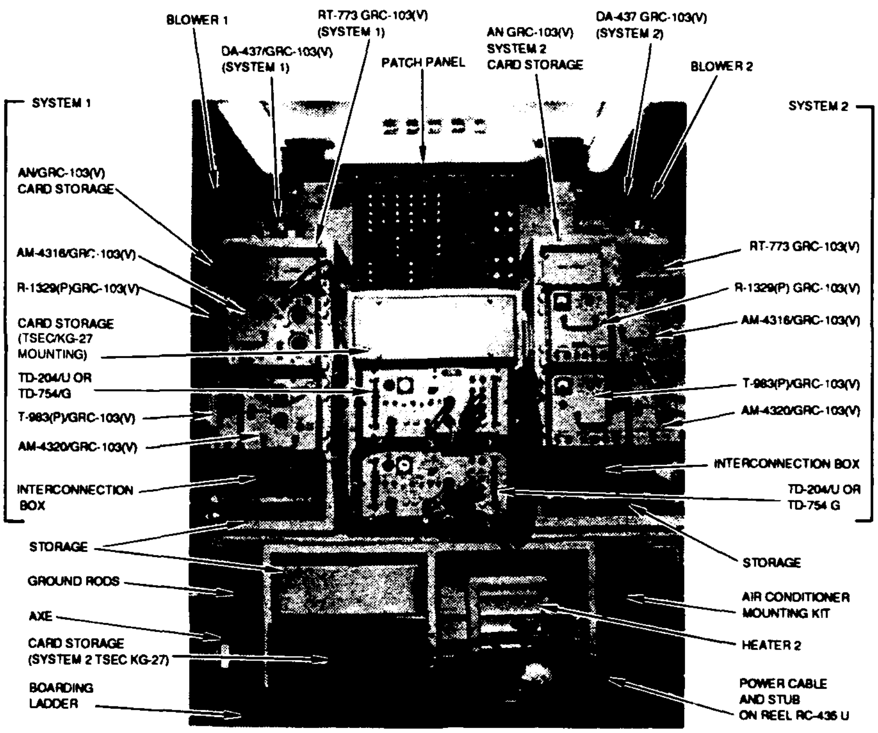

Fig. 2.1—AN/TRC-145 Radio Terminal

(most commonly) two 12-channel terminals, each consisting of a radio transmitter and receiver, signal converter, security device, and two multiplexers (devices that combine or decouple two or more signals on a single channel) (Department of the Army, 1977b).

Members of MOS 31M are ordinarily assigned to all echelons. The numbers of active-duty enlisted MOS 31M personnel and accessions in FY 1989 are shown in Table 2.1.

Table 2.1

Operating Strength and Accessions, MOS 31M

Operating strength, FY 1989 (by pay grade and skill level)	
E1–E3 (level 1)	1515
E4 (level 1)	2671
E5 (level 2)	1361
E6 (level 3)	1071
Total	6618
Accessions	
FY 1989 (actual)	1198
FY 1990 (estimated)	1280

SOURCE: U.S. Total Army Personnel Command (1989).

REACTIVE ELECTRONIC EQUIPMENT SIMULATOR

Description

The REES is a computer-*controlled*, *high-fidelity* simulation facility located at Fort Gordon, Georgia, the site of the U.S. Army Signal Center and School, where Advanced Individual Training (AIT) is provided for communications and electronics occupations. The REES contains four communications *nodes* representing separate field locations (*signal centers*). Each node contains seven separate sets of the related communications equipment (assemblages) that would be operated in the field by the appropriate signal specialists.[5]

The REES also contains a central instructor console from which actions can be monitored and the system controlled. Each assemblage can operate independently or as one component of an integrated tactical communications network, depending on the *configuration* of assemblages and nodes established by the system controller. Alternative communications systems are simulated through use of specific nodes and assemblages and the links established between them. Networks can be established to represent alternative tactical environments. For example, a corps could be represented using the equipment in a node that would be found customarily in a corps signal center. Meanwhile, division and lower-echelon signal centers could be

[5]Each node contains one AN/TRC-145 and one AN/TRC-151 (operated by MOS 31M), one AN/TRC-138 (operated by MOS 31Q, Tactical Satellite/Microwave Systems Operator), three AN/TCC-73s (operated by MOS 31M or MOS 31Q), and one AN/TSQ-84 (operated by MOS 31N, Tactical Circuit Controller).

represented in other nodes. Further, different nodes could be connected to each other in different ways (e.g., using radio or cable), as doctrine would suggest is appropriate in the battle situation.

A key feature of the REES is that it contains the actual faceplates, panels, and switches of the real communications equipment. The operators use the same equipment and follow the same procedures they would in a shelter in a true field environment. Rather than transmitting actual signals, however, the assemblages "communicate" through a master computer. The master computer evaluates individual actions taken on each piece of equipment and causes the system to respond appropriately to those actions. It also extracts and saves data, which we used as the basis for our analyses.

Advantages for Performance Assessment

The REES offers several important advantages for assessing group performance in operating and troubleshooting communications systems:

- It provides a realistic simulation of communications assemblages and networks in which task success implies the availability of division/corps command-and-control facilities.
- Its simulator can be used to provide an objective test that can be administered consistently and under controlled conditions across test administrations.
- Test difficulty can be manipulated, through test conditions and standards, to represent field tasks of varying difficulty and to provide desirable statistical variation in outcomes.
- Its computer can provide considerable detail on the performance of individuals and groups, producing data whose likelihood of error is less than that of similar measures made without computer assistance.

Altogether, these advantages indicate that the REES can provide an objective test that is likely to produce reliable and valid measures of individual and team proficiency.[6] Because the REES uses the actual faceplates of the equipment and the tasks are identical to those per-

[6]*Reliable* here means consistent from test administration to test administration. *Valid* means accurate with respect to true success or failure at communications system operation and troubleshooting. If the protocol is followed consistently, the conditions of testing and measurement should not vary from administration to administration.

formed with field equipment, a soldier's ability to perform in the REES should be highly related to his ability to perform in a shelter outside the REES. Indeed, given the data collection capabilities of the REES (not present in the tactical equipment), testing in the REES is preferable to a test involving actual equipment, which would of necessity require some subjective assessment of performance.

In addition, experience has shown that the REES can provide a realistic simulation of communications networks that might be established in combat. The assemblages look and act like the real thing (they have "face validity"), their operation procedures correspond to those prescribed in relevant technical manuals, and the simulations conform with established communications doctrine (Gould, 1981). For example, a typical network configuration employs the four nodes in the REES to simulate a communications network between one armored division, two infantry divisions, and one corps headquarters.

TEST PROCEDURES

Based on earlier RAND experience using the REES for individual performance assessment (Winkler and Polich, 1990) and in conjunction with subject-matter experts at the Army Signal Center, we designed and conducted tests of communications system operation and troubleshooting. These tests examined the effect of personnel aptitude (AFQT score) on the performance of groups of communications operators as they interacted to operate systems and isolate faults in a command-and-control communications network that could be established in wartime.

Simulated Configuration

Figure 2.2 shows the simulated network configuration, which consisted of four assemblages: two *terminals* and two *relays*. The terminals represented the assemblages that might be found at nodes such as the division or corps main command post (terminal A) and division artillery (terminal B). They were connected by a relay system at intermediate nodes, where the signals from both terminals were received and retransmitted. Each terminal managed two separate 12-channel systems.[7] The relay was used as a "repeater" to forward signals in one 12-channel system from one terminal to the other. In

[7]In an operational system, each channel would in turn contain several separate lines of communication, including voice, code, data, and so forth.

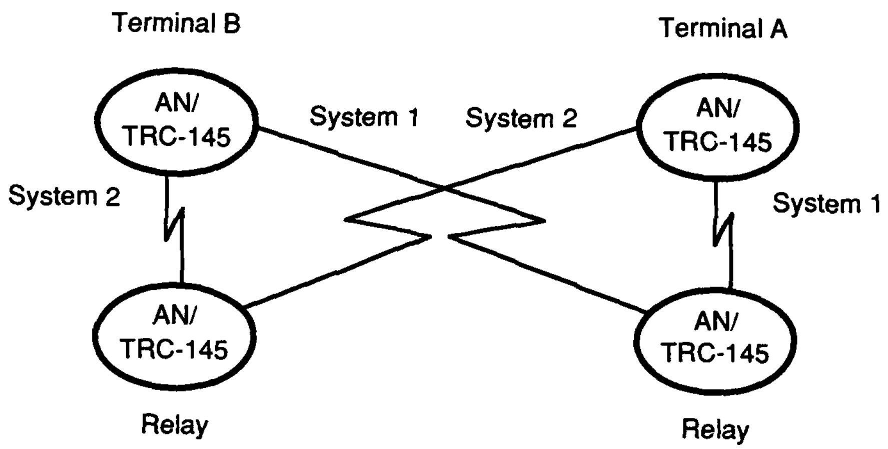

Fig. 2.2—Test Configuration

addition, the nodes were connected by a separate, single line of communication (an "order wire") to allow operators to communicate with each other. The test in the REES employed groups of three operators: two persons assigned to the terminal positions and one person assigned to a relay position.[8]

Equipment

Each node in the REES contains an AN/TRC-145 and an AN/TRC-151, which are the principal pieces of radio equipment used by 31Ms. We tested 31Ms on the AN/TRC-145 because, according to subject-matter experts at the Signal School, it is used more widely in the field than is the AN/TRC-151, which is used only at corps and higher echelons. This decision allowed us to test a larger number of unit personnel on applicable equipment. Also, when our research was being performed, the AN/TRC-151 was receiving only minimal attention during AIT. Thus, the AN/TRC-145 offered greater advantages for testing purposes.

[8]Through most of the test, the relay was placed in the first 12-channel system (system 1). In part of the test of AIT graduates, conducted to provide additional information for the Signal Center about the relationship between relay and terminal performance, the relay was placed in the second 12-channel system (system 2). The unmanned relay was placed in "override" position by the REES computer, meaning that it was, in effect, transparent to other users in the system.

Examinees

The test involved members of MOS 31M drawn from two separate populations: graduates of AIT and personnel from four active-duty signal battalions. The testing was carried out from September 1988 through April 1989.[9] Operators were tested in groups of three and were preassigned to groups and nodes within each population using a random assignment method. We had determined that such a method would be likely to produce wide variation in the "quality" of teams, thus mirroring the types of teams the Army would have in the field in a conflict. To investigate whether random assignment would provide a suitable distribution of individual and group quality for a team test, we first examined the underlying distribution of aptitude (AFQT score) among FY 1988 31M non-prior service accessions. Table 2.2 shows these figures for active-duty accessions.

How are these individuals likely to be assigned to communications nodes? If the commanders' information and control were perfect, the most capable or experienced individuals would most likely be placed in strategically important locations so that individual nodes and the network would be formed to maximize system performance. We suspect, however, that the exigencies of battle would preclude any such systematic assignment. In an actual tactical situation (during armed conflict), individual specialists would be spread out over an extended geographic area. Each small group of communications operators would move frequently as the combat units maneuvered across the battlefield area. Over time, a given operator could find himself in dif-

Table 2.2

Distribution of AFQT Scores in MOS 31M: FY 1988 Active-Duty Accessions

AFQT Score by Mean Percentile and Category	Percentage of 31Ms
65–99, category I, II	38
50–64, category IIIA	29
10–49, categories IIIB, IV	33

SOURCE: Data provided by Defense Manpower Data Center.

[9]The numbers of groups tested in system operation and troubleshooting differed. See Secs. 3 and 5 for details.

ferent locations and/or performing different network functions (e.g., terminal one day, relay the next). The specific role played by any single specialist would probably be determined by a process that would be much more random and less predictable than that used in peacetime exercises.

If a random process governs operator assignment, we can use the binomial function to estimate the distribution of quality that would be observed in groups drawn from a large sample. Based on the binomial process and the fact that 67 percent of the underlying population is in categories I, II, and IIIA, randomly drawn groups of three 31Ms are likely to have the following characteristics: 30 percent will have all three members in AFQT categories I through IIIA; 44 percent will have two such members; 22 percent will have just one such member; and 4 percent will have none.

We concluded that a random assignment mechanism should be sufficient to ensure broad variation in the overall aptitude levels of groups of operators. To carry out the assignments, we obtained lists of potential examinees in advance of the testing and allocated individuals at random to teams of three and to REES nodes within teams. This procedure was followed for forming groups of AIT graduates as each successive class of trainees completed the course, and for forming groups of personnel from active-duty signal battalions during designated test periods. These groups were then tested on the two primary tasks, system operation and troubleshooting.

3. SYSTEM OPERATION

This section focuses on our test of system operation performance, in which groups of three persons were required to install and operate their equipment to establish a working tactical communications system. We first describe the details of the testing procedures and performance measures; we then describe the estimated effects of AFQT score on group success.

TESTING PROCEDURES

We preassigned individual soldiers to three-person groups and specific assemblages at random within each of the test populations (i.e., among graduating classes of AIT students and within active-duty signal battalions). Each group received an initial briefing to describe the tactical situation and specific REES procedures and inform the group that its initial task would be to operate a radio system using the AN/TRC-145.[1] Individuals were instructed to install their assemblage as either a radio terminal or a relay, depending on their preassigned node locations. They were then directed to their nodes within the REES facility, where they received further written instructions regarding their "mission" (i.e., "cut sheets" assigning frequencies on which to transmit and receive). After the console operator had initialized the system, the group members entered identifiers via the computer keyboard attached to each assemblage. They were then instructed to begin the system operation task, following procedures in the REES performance guide (U.S. Army Signal Center, 1983), which adheres to standards in the relevant technical manuals. Upon completion of the task, team members depressed a "task-stop" button on their computer keyboards.

The objective of this task was to establish communication between the terminal positions in the network. Performing this task involved individual installation of the equipment followed by group interaction to establish communication between nodes. As the team began its work, each soldier had to correctly interconnect the individual components in his assemblage, preset and perform operational checks of

[1]One purpose of the briefing was to present the tactical environment and operating procedures to the operators just as would be done in an actual unit prior to deployment. Another objective was to ensure that the soldiers were equally familiar with the REES operating procedures.

the components, and align and adjust his own equipment. The soldiers then sought to establish initial communication with each other. The soldier at the relay installed his assemblage and then made contact with each terminal operator. Next, the terminal operators established contact with each other on the first 12-channel system through the relay, aligning and adjusting their equipment as needed to ensure adequate communication. The terminal operators then repeated these procedures on the second 12-channel system.[2] All operators had the ability to use the order wire as a back-channel for communicating and resolving difficulties. Seventy minutes were allotted for this task, consistent with Army standards established for unit performance.[3]

Measurement of Group Performance

Our primary performance measure was whether the communications system became operational (i.e., whether both 12-channel systems functioned) within the allotted time. We obtained measures of performance from two sources: global ratings of success provided by signal experts, and detailed records of operator actions maintained by the REES computer. The global ratings were reported in writing by a test supervisor and other observer/controllers (O/Cs), who were unaware of the AFQT categories of the examinees.[4] At the conclusion of the time allotted for the task, the test supervisor judged the overall system as usable or unusable ("go" or "no-go") based on observation of operator performance and, where necessary, a speech test on each of the systems.

As validation for these judgments, we relied on the data base maintained by the REES computer. The REES computer data showed, for

[2]There are two phases to operation of each system. In the initial phase, the system is installed in *nonsecure* mode, i.e., without the communications security (encryption) system. Then, the system is realigned using the communications security system.

[3]The ARTEP (Army Training and Evaluation Program) allows 120 minutes to install and operate two 12-channel radio systems using the AN/TRC-145. These standards, however, pertain to three-man crews assigned to a single shelter, and they include additional tasks—e.g., selecting an appropriate site, preparing and camouflaging the shelter, maintaining security, preparing and operating a generator, and installing an antenna (which itself requires three persons and 30 minutes). The consensus of signal subject-matter experts, including the general officer serving as assistant commandant of the Signal Center, was that 70 minutes should be more than sufficient for an individual 31M to operate the multichannel equipment.

[4]The test supervisor was a retired 31M sergeant major who was hired, trained, and supervised by RAND. The other O/Cs included the senior NCO operator of the REES computer (the console operator) and additional civilian and military assessors who observed examinees at work in the nodes.

each individual assemblage, whether errors had been made during execution of the task.[5] The data did not, however, directly indicate the final status of the communications system as a whole or provide other measures of overall group performance. To derive such an overall measure, we linked the individual records indicating the status of each assemblage at the last task-stop to determine whether the assemblages were operational. The system as a whole was inferred to be operational when all assemblages were operational. Among the AIT graduates, scoring of system operation based on the overall judgment of the test supervisor agreed with the computerized measure for 237 of the 240 groups (99 percent).

In our final analysis, however, we used the measure of performance provided by the test supervisor, because the computerized validation was not available for unit members. In units, standard operating procedures often differ from the technical-manual procedures embodied in the REES software (e.g., unit members often omit nonsecure operation and use other procedures found to be expedient in the field).[6] During the testing of unit personnel, participants were allowed to follow the standard operating procedures of their units, although the time standard remained unchanged. As with the AIT graduates, the O/Cs carefully monitored the examinees, and the test supervisor judged, at the conclusion of testing, whether the system was usable. However, the REES computer software recorded alternative procedures followed by unit members as errors and thus could not be used to conduct a detailed validation for unit members.

[5]Specifically, at each depression of the task-stop button at the assemblage, the computer evaluates the steps taken and, for any detected errors in the setup or switch actions, illuminates a red light on the computer keyboard and at the operator's console. In addition, the computer generates a record that shows at each task-stop whether the task was performed correctly or incorrectly, cumulative time on the task, and other information. Interpretation of the computerized data is complicated, however, because the computer may indicate an error for various reasons. An indication of *incomplete* may be generated if the assemblage simply does not work (e.g., if cables are incorrectly patched). It may also be generated for more minor mistakes, such as performing steps outlined in the performance guide out of sequence. According to the signal subject-matter experts, the latter indications of error represent "false negatives." Systems could be operated using short cuts or combinations of steps other than those allowed by the computerized scoring rules.

[6]Some individuals may have had difficulty because they were not currently working with the AN/TRC-145 or in MOS 31M, for which they were officially qualified. This problem was not widespread, however. In questionnaires collected from all unit members, most reported that their current duty MOS was 31M (92 percent), and most said that they were currently working with an AN/TRC-145 or equivalent (68 percent).

Basic Data Distributions

Test Results. By the conclusion of testing, usable data had been collected from 720 AIT graduates and 252 unit members who had been properly configured into three-person groups.[7] The performance of operator groups for each of the test populations is shown in Table 3.1. The system was judged as operational within the time limit for 61 percent of the groups of unit members and for 41 percent of the groups of AIT graduates. The remaining groups were unable to correctly operate the communications system within the allotted time.

Test Participants. The characteristics of the individual personnel upon whom the analyses were based are shown in Table 3.2. The figures were derived from records of the Defense Manpower Data Center (DMDC), which maintains data on written and physical tests for enlistment qualification.[8] The groups contain slightly different AFQT score distributions, with the AIT graduates containing relatively fewer category IIIB and IV personnel than were found in the unit sample.[9] Neither group, however, contained a sizable number of personnel in AFQT category IV.

Table 3.2 also shows that the populations of AIT graduates and unit members differed from each other in background and experience. The unit members were older, and about a fifth of the unit members were

Table 3.1

Group Performance at System Operation (in percent)

	Test Group		
Performance	AIT Graduates (240 groups)	Unit Members (84 groups)	Total Sample (324 groups)
Operated successfully within time limit	40.8	60.7	46.0
Not operated successfully	59.2	39.3	54.0

[7]Four AIT groups (12 members) were discarded from the analysis because we could not determine the AFQT score of one or more members. An additional trio of unit members was discarded because one member had participated earlier as an AIT graduate.

[8]In the few cases for which no DMDC data existed, data were obtained from Army-maintained personnel files.

[9]AFQT categories are defined by percentile scores normed on the U.S. youth population: category I, percentiles 93–99; category II, 65–92; category IIIA, 50–64; category IIIB, 31–49; category IV, 10–30. Category V persons, percentiles 1–9, are excluded by law from military service.

Table 3.2

Characteristics of Individual 31Ms Tested in System Operation

Characteristic	AIT Graduates	Active-Duty Personnel
AFQT category (percent)		
I	2.1	2.0
II	34.0	31.7
IIIA	24.4	21.4
IIIB	38.1	33.3
IV	1.4	11.5
AFQT score, mean (s.d.)	58.0 (16.8)	54.3 (20.6)
Age		
18–19	51.9	2.4
20–21	24.4	30.8
22–23	10.4	26.3
24–25	5.9	13.0
26 or older	7.4	27.5
Age, mean (s.d.)	20.6 (3.0)	24.1 (4.1)
Sex (percent)		
Male	81.8	90.9
Female	18.2	9.0
Race (percent)		
White	66.8	55.5
Nonwhite	33.2	44.5
Education (percent)		
High school graduate	94.0	88.1
Nongraduate	6.0	11.9
Component (percent)		
Active	71 [illegible]	100.0
Guard or Reserve	28.2	0.0
Pay grade (percent)		
E1	76.8	0.8
E2	9.8	3.6
E3	11.6	26.2
E4	1.8	46.8
E5	0.0	21.0
E6	0.0	1.6

NOTE: Based on 720 AIT graduates and 252 active-duty personnel.

above the initial skill level conferred upon AIT graduation (as indicated by a grade of E5 or higher). The AIT graduates, moreover, included a number of Reserve component personnel, who were generally non-prior service enlistees on temporary active-duty status during training. We controlled for these differences in group characteristics in our analyses.

Distribution of Group Aptitude. Table 3.3 shows the distribution of aptitude across the groups. The final sample of participants was configured into 240 three-person groups of AIT graduates and 84 trios of unit members. The average AFQT score of the three group members, which we used as a summary measure of group aptitude, fell within category IIIA about 50 percent of the time for both populations. Another way to characterize overall group aptitude is to count the number of "high-scoring" members (i.e., with AFQT scores in categories I through IIIA). As can be seen in Table 3.3, the modal (most common) team contained one member from the lower AFQT categories and two members from the higher AFQT categories.

The distributions of average AFQT score and number of high-scoring members indicate substantial variability in group aptitude in both test populations across AFQT categories, except within category I and

Table 3.3

Group Aptitude Distribution for System Operation Test (in percent)

Group AFQT Composition		AIT Graduates	Active-Duty Personnel
Average AFQT score distribution by category and mean percentile			
I,	93–99	0.0	0.0
II,	79–92	2.5	1.2
II,	65–78	22.5	15.5
IIIA,	57–64	27.5	21.4
IIIA,	50–56	28.8	28.6
IIIB,	40–49	16.6	23.8
IIIB,	31–39	2.1	8.3
IV,	10–30	0.0	1.2
Number of members in AFQT categories			
Three I–IIIA personnel		22.9	15.5
Two I–IIIA personnel, one IIIB–IV		41.7	45.2
Two or more IIIB–IV personnel		35.4	39.3

NOTE: Based on 240 three-person groups of AIT graduates and 84 three-person groups of active-duty personnel.

category IV. Thus, there should have been enough operator groups in the higher and lower categories to allow us to predict the effect of raising or lowering AFQT scores across a wide range of scores.

ANALYSES OF GROUP APTITUDE AND PERFORMANCE

Analytic Method

To assess the effects of aptitude on performance in operating communications systems, we used regression models that included aptitude and additional personnel characteristics as predictors of performance. Group performance in operating the system—completing the task within the time standard—was treated as success or failure, a binary (0/1) measure. Therefore, we used logistic regression to estimate statistical relationships.[10] Using this model, we examined the effect of alternative measures of group aptitude while controlling for other group differences. This approach allowed us to estimate the effects of AFQT score on group performance more precisely ("net" of other differences among groups) while providing a basis for predicting how performance would differ as group aptitude changed over AFQT categories. Predictor variables in the model included the following:

- Age of group members (average).
- Variables representing the number of group members that were male, white, and high school graduates (each coded 0 through 3).
- A dummy variable indicating whether the test group was composed of unit members (coded 1) or AIT graduates (coded 0).[11]
- The number of group members currently using the AN/TRC-145 in their regular job, a measure of equipment familiarity (coded 3 for AIT graduates and 0 through 3 for unit members).[12]

[10]Logistic regression is the appropriate functional form for conducting regression analysis of binary (0/1) variables. The form of the equation is $y = 1/[1+EXP(-bx)]$, where y is the outcome variable, EXP is the exponential function (base e), x is a vector of independent variables, and b is a vector of their coefficients. This form permits interpretation of the outcomes as the probability of y (e.g., success) given the vector of independent variables, x.

[11]In preliminary analyses, we examined alternative measures of job experience for unit personnel obtained from a questionnaire administered to unit members. The purpose of this instrument was to provide measures of experience for unit members, including information on current work assignments and recent job-related experiences. We found, however, that these measures were substantially intercorrelated, precluding their concurrent use in the analysis. Among measures of job-related experience, unit membership proved most robust.

[12]This measure was taken from the questionnaire filled out by unit personnel.

- A dummy variable indicating whether the test group contained any members drawn from the Reserve component (coded 1 versus 0).

In this model, we also include a dummy variable to control for alternative test regimens in which the relay was located in either system 1 or system 2 (see Sec. 2). This variable was coded 1 for those AIT groups in which the relay was located in system 2, and 0 for all remaining groups. Means and standard deviations of the predictor variables are shown in Table 3.4.

Testing Possible Group Effects

Initially we assumed that the most straightforward measure of group aptitude was the average, or mean, AFQT score of the three group members. This formulation seemed most consistent with the purpose of our analysis, which was to estimate the effects on group perfor-

Table 3.4

Predictor Variables Used in the Models, by Test Population

	AIT Graduates		Unit Members		Combined Sample	
Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.
Successful operation (no/yes)	0.41	0.49	0.61	0.49	0.46	0.50
AFQT score (average)	57.95	9.60	54.28	11.02	57.06	10.06
AFQT score distribution (number in categories I–IIIA)	1.82	0.85	1.65	0.87	1.78	0.86
Number of members using equipment	3.00	0.00	1.94	0.86	2.72	0.64
Age (average)	20.60	1.94	24.14	2.56	21.51	2.63
Education (number of high school graduates)	2.80	0.46	2.56	0.55	2.74	0.49
Sex (number male)	2.45	0.72	2.73	0.50	2.52	0.68
Race (number white)	2.00	0.80	1.63	0.86	1.90	0.83
Reservists present in group (no/yes)	0.57	0.50	0.00	0.00	0.42	0.49
Relay position (system 2)	0.48	0.50	0.00	0.00	0.36	0.48
Number of groups	240		84		324	

mance of lowering or raising the aggregate level of aptitude among communications operators (e.g., from category IIIA to category IIIB). The results of our initial analysis, shown in Table 3.5, indicate that the group's average AFQT score is significantly related to the group's ability to operate communications systems successfully when measures of demographic background, education, and military experience are controlled. Groups with higher average AFQT scores perform better than do groups with lower average AFQT scores.[13]

Table 3.5

Regression of System Operation and Average Group AFQT Score

Variable	Coefficient	Standard Error	Chi-Square
Average group AFQT score	0.041	0.013	10.04*
Test population (unit members)	1.766	0.529	11.14*
Number of members using equipment	0.440	0.282	2.44
Average age of operators	–0.110	0.058	3.65
Number of high school graduates	0.034	0.252	0.02
Any Reservists in group	0.255	0.287	0.79
Number of males	0.134	0.180	0.55
Number of whites	0.084	0.153	0.30
Test regimen (relay in system 2)	–0.479	0.276	3.01
Intercept	–2.338	1.930	1.47

NOTE: Logit model used, based on 323 groups with complete data. Model significant at $p < .001$ (chi-square = 32.49; $-2 \log L = 413.02$). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

[13]The AFQT coefficient shown in Table 3.5 is unbiased for the members of MOS 31M examined in the study. The magnitude of the AFQT score effect for enlistees in general may differ from what is shown if soldiers who enlist in MOS 31M differ from other enlistees according to other unmeasured factors. For example, enlistees may self-select for this MOS, or they may be influenced by military job counselors to choose this specialty because they have other attributes believed to improve their performance as communications system operators. Such selection would restrict the range of observed outcomes. When this selection effect is taken into account via a Heckman correction for selectivity (Greene, 1990, p. 744), the magnitude of the AFQT coefficient changes very little, increasing by approximately 9 percent. The correction term itself was not a statistically significant predictor of group performance. We therefore chose to estimate the effects of AFQT score on performance from our empirical sample in the remaining analyses.

Alternative Measures of Group Aptitude. Our initial analyses used average AFQT score as a measure of group aptitude, but alternative specifications for aptitude that represent competing theories of group dynamics could be considered. The influence of aptitude on group performance could depend on the type of task performed by the group or the attributes of individual group members. Moreover, dynamic properties of group interaction could affect performance beyond the scope of individual members' abilities. To examine such possibilities, we conducted sensitivity analyses including alternative measures of individual and group aptitude. In general, these analyses indicated that our model using average group AFQT score was robust for a wide range of alternative specifications and the addition of other variables. Below, we describe alternative models we examined. Regression tables showing the results of major alternative analyses are contained in App. A.

Task Characteristics. Steiner's discussion (1972) of task demands and productivity has been extremely influential in analyzing group performance in a variety of settings. Steiner argues that task requirements are key determinants of group performance and that tasks could be classified based on the individual contributions needed to achieve a successful group outcome. According to his taxonomy, major distinctions include disjunctive, conjunctive, and additive tasks.[14]

In a *disjunctive* task, for example, the group succeeds if a single member succeeds. Consider, for example, a competition in which student teams attempt to solve a math problem. The first team to solve the problem wins. In such a case, the team most likely to win will contain an individual with the resources needed to win the competition. In general, the team's performance will be determined by its most able member. In a *conjunctive* task, each group member must perform the same function, and all must do well for the group to succeed. An example is the speed with which a team of mountain climbers reach a summit. In this situation, group performance is determined by its least able member. Finally, in an *additive* task, group performance depends on all individual performances. Thus, the outcome is a combination of individual outcomes and does not depend on a specific group member. An example is a bowling team whose team score combines individual performances.

---

[14]See Kahan et al. (1984) for a more extended discussion of the theoretical literature on the link between group behavior and military unit performance.

This discussion implies that the role of aptitude in determining group performance in communications system operation may depend on whether the task is disjunctive, conjunctive, or additive. Alternative formulations of our model can be specified according to each of these possibilities. For example, if the task is seen as conjunctive, the performance of one "weak link" should cause the group to fail. In this case, to analyze group performance, aptitude might be represented as the minimum AFQT score among the group members. If a capable "strong leader" drives group performance, then aptitude might be represented as the maximum AFQT score among the group members. Finally, if successful performance by all of the operators is needed for group success, then the task could be viewed as additive, and group aptitude would be best represented as a combination of individual aptitudes.

Given these possibilities, we viewed the task as primarily additive in nature. Key task characteristics included the high degree of coordination among operators during task performance and the ability of operators to communicate and offer assistance to each other (via the "order wire").[15] Among these, task coordination could be especially important, since all members were required to contribute (O'Brien and Owens, 1969). Thus, we initially preferred a model that combined individual aptitudes (i.e., average AFQT score). To gain empirical insight, however, we examined logistic regression models that represented aptitude as either the effects of a strong leader (maximum AFQT score) or a weak link (minimum AFQT score). Our results (shown in App. A, Tables A.1 and A.2) demonstrate that these models were not as good as the model using average AFQT score. The chi-square of the average AFQT score model, a measure of goodness of fit, was larger than the chi-squares of these two alternative models.[16] Another means of comparing the models was provided by the –2 log likelihood ratio, which indicates the amount of variance unexplained by the model. Compared to the other two models, the average AFQT score model left less variance unexplained.[17,18] Thus, we saw no

[15]Our O/Cs noted that groups generally were highly motivated to succeed. Members communicated frequently with others using the order wire and assisted each other in resolving problems when they occurred.

[16]The chi-square of the average AFQT score model was 32.49. The chi-squares of the maximum and minimum AFQT score models were, respectively, 27.15 and 27.36.

[17]For the average, maximum, and minimum AFQT score models, the results were $-2 \log L = 413.02, 418.37,$ and 418.16, respectively.

[18]In addition, we compared the simultaneous effects of average AFQT score and maximum *or* minimum AFQT score in two additional models. These models, shown in App. A, Tables A.3 and A.4, indicate that average AFQT score is a stronger predictor of group performance than either the maximum or minimum AFQT score within the

reason to prefer a formulation in which group performance was determined by the aptitude of one specific group member.[19]

Individual Contributions. Assuming that group performance can be analyzed as a combination of individual aptitudes, alternative specifications can be considered for combining aptitude measures. Our preference was to represent the aptitudes as a group average, since this approach appeared most relevant to the policy problem of changing the aptitude profile across the MOS. An alternative, however, was to represent aptitude in the group by using the individual AFQT scores of the group members in place of the group mean.

The results of this analysis, given in Table 3.6, show that group performance improves as the AFQT scores of the three members increase. The relationships are all positive, and two of the AFQT coefficients surpass the conventional significance level (.05). In addition, the total of the individual coefficients is the same magnitude as the coefficient when the measures are combined as average AFQT score (Table 3.5). Thus, the model suggests that the effects of individual AFQT scores are additive for this task, but the average AFQT model represents the phenomena more parsimoniously. In fact, when we formally compared the predictive power of these two models, we found that the difference was trivial and far from statistically significant (chi-square = 0.69, n.s.).[20] We concluded that the aptitude of group members performing this task could be specified as average AFQT score for our analysis.

Group Productivity Effects. As described thus far, our analysis suggests that individual aptitudes affect group performance in an

---

group when one of the latter measures is included along with the average AFQT score. The effect of average AFQT score remains statistically significant, whereas the effect of maximum or minimum AFQT score is not significant.

[19]Another possibility that we considered but rejected based on empirical evidence was the effect on the group outcome of single operators in specific positions. Here, we examined whether the group outcome was determined by the aptitude of the member in terminal A, the relay, or terminal B. Compared to each of these three cases, the model using the average AFQT score of the three members provided a better fit with higher chi-square and smaller –2 log likelihood values. See App. A, Tables A.5, A.6, and A.7, for details.

[20]We evaluated these alternative models by comparing the ratio of the log likelihoods of the baseline model using average AFQT score and the model containing the three operators' AFQT scores (Maddala, 1988, pp. 84–85). This ratio has a chi-square distribution with degrees of freedom equal to the number of additional coefficients. Here, the small improvement in predictive power afforded by the individual AFQT model (–2 log L = 412.29 versus 413.02 for the model using average AFQT score) did not warrant the two additional degrees of freedom required by this model.

Table 3.6

Regression of System Operation and Individual AFQT Score

Variable	Coefficient	Standard Error	Chi-Square
AFQT of terminal A operator	0.017	0.007	6.21*
AFQT of relay operator	0.009	0.007	1.75
AFQT of terminal B operator	0.015	0.007	4.91*
Test population (unit members)	1.799	0.532	11.41*
Number of members using equipment	0.434	0.283	2.34
Average age of operators	–0.112	0.058	3.77
Number of high school graduates	0.032	0.253	0.02
Any Reservists in group	0.264	0.288	0.84
Number of males	0.127	0.180	0.49
Number of whites	0.079	0.153	0.27
Test regimen (relay in system 2)	–0.472	0.277	2.90
Intercept	–2.261	1.936	1.36

NOTE: Logit model used, based on 323 groups with complete data. Model significant at $p < .001$ (chi-square = 33.22; $-2 \log L = 412.29$). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

additive fashion and assumes that the effect of aptitude on performance is linear. One might argue, however, that group interaction could cause performance to exceed that expected by combining attributes of group members (e.g., Hackman and Morris, 1983). Such a "compositional" effect is said to occur, for instance, if the result of adding a high-ability person depends on the value of the other persons in the group. Thus, for example, there would be a compositional effect if increasing the aptitude of one of the terminal operators made more of a difference if the other terminal operator's AFQT score was high rather than low. Tziner and Eden (1985) describe some evidence for the compositional effects of ability on the performance of Israeli tank crews as rated by supervisors.[21]

To test for such phenomena, we examined additional models of group performance. Specifically, we looked at models that included a quadratic term (the square of the average AFQT score, in addition to the average AFQT score) to test the linearity of the relationship be-

[21]But a second study found that team performance in an interdependent task is not likely to be affected in a nonadditive manner by team composition (Tziner, 1988).

tween AFQT score and performance, as well as some additional functional forms designed to examine the linearity of the model.[22] We also examined models that included terms for interaction among the individuals' AFQT scores. None of the additional terms (shown in App. A, Tables A.8 and A.9) proved statistically significant. The predictive power of these models was not significantly greater than that of the baseline model using average AFQT score, based on tests of the log likelihood ratios.[23] Thus, the effects of member AFQT score on group performance appear to be linear, and there is no evidence that member interaction improves performance beyond that predicted by the individuals' aptitudes.

As further evidence that the effects of AFQT score are additive among group members, we considered the value of having "just one smart person" or multiple "smart people" in a group. Our analyses suggested that each individual contributes to group success, implying that the more high-scoring individuals, the better. This finding is explicitly illustrated by the analysis results in Table 3.7. Here, success or failure of the group was modeled as a function of the *number* of high-aptitude persons in the group.[24] The results show that the probability of successful system operation depends on the number of high-aptitude personnel in the group. As the number of high-aptitude members increases, the group is more likely to succeed.

Differences Among Test Populations. The analysis results reported thus far led us to prefer our initial model using average AFQT score (Table 3.5) to alternative models for analyzing the effects of AFQT score on the ability of signal operators to operate communica-

[22]For example, there is some reason to believe that the effect of AFQT score would not be linear, but rather linear in logs. That is, it is reasonable to hypothesize that the effect of AFQT score is greatest among the middle range of scores and that changes in AFQT score at the extreme upper and lower end of the AFQT scale would make little difference. Entering group AFQT score in log form (as the mean of the logs of the AFQT scores of the three soldiers) produced results that differed very little from those produced using the actual AFQT score, however. We also examined multiplicative effects of AFQT score (measured as the log of each individual AFQT score). None of the alternative models provided significantly improved predictability or added interpretability compared to the baseline model using mean group AFQT score. We therefore concluded that, at least in the range of AFQT scores of soldiers we tested, the effect of AFQT score appeared to be linear and that by entering AFQT measures linearly, we would avoid the additional complexities of interpreting the results that would be introduced by log formulations.

[23]The chi-squares for the quadratic and member interaction models, compared to the chi-square for the baseline model using average AFQT score, were 0.32 (n.s.) and 0.19 (n.s.), respectively, based on log likelihood ratio tests.

[24]A given person is regarded as *high-aptitude* if his AFQT percentile score is 50 or above—a definition commonly used by the military services and Congress.

Table 3.7

Regression of System Operation and Number of High-Aptitude Personnel

Variable	Coefficient	Standard Error	Chi-Square
Number of members in categories I–IIIA	0.346	0.146	5.57*
Test population (unit members)	1.605	0.522	9.47*
Number of members using equipment	0.436	0.278	2.46
Average age of operators	–0.091	0.057	2.58
Number of high school graduates	0.022	0.250	0.01
Any Reservists in group	0.258	0.285	0.82
Number of males	0.111	0.180	0.38
Number of whites	0.155	0.148	1.10
Test regimen (relay in system 2)	–0.507	0.275	3.40
Intercept	–0.998	1.826	0.30

NOTE: Logit model used, based on 323 groups with complete data. Model significant at $p < .01$ (chi-square = 27.6; $-2 \log L = 417.82$). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

tions systems. An analytical issue remained, however, regarding our pooling of results for AIT graduates and unit members.

Our models showed consistently that groups of unit members performed better than groups of AIT graduates. We deemphasized direct comparisons of performance between unit members and AIT graduates because unit members installed systems according to their unit operating procedures, whereas AIT students followed the practices adhering to the technical manual. There was a remaining issue, however: whether the procedural and performance differences for these populations called for separate regressions for modeling the effects of AFQT score even though the standard was the same in both cases—i.e., having test supervisors judge whether a system was usable.

We conducted a number of additional analyses to explore this issue and concluded that pooling was appropriate. We first tested whether the variances of our outcome measure (successful system operation) differed between the populations. No evidence was found that the variances of the two groups differed ($F = 1.01$ with 239 and 84 degrees of freedom). Thus, it is unlikely that across-group variation biased the regression coefficients. In addition, when we estimated separate

models for AIT graduates and unit members while holding constant common measures of background and experience (average age and number of whites, males, and high school graduates), we found that the AFQT coefficients were nearly identical (see App. A, Tables A.10 and A.11).[25] Finally, as a further test of the need for separate models for each population, we examined a pooled model containing interaction terms between test population and the other predictor variables (see App. A, Table A.12). None of the interaction terms was statistically significant, and their inclusion did not improve the model's predictive power.[26] We thus concluded that observations from both test populations could be pooled and that the inclusion of a single term for group affiliation adequately captured differences between these populations.

PREDICTIONS FROM THE MODELS

In an analysis of the type we conducted, predictions from the models are often used to interpret results. Such predictions can be used to compare subgroups of interest, e.g., differences in performance that would be expected from groups whose average AFQT score falls within different AFQT categories, such as category IIIA versus IIIB.[27] Because AIT graduates and unit members differed in test procedures and performance, we examined the relationships between AFQT score and performance separately for these populations. Below, we first discuss the predictions of performance we arrived at based on our model using average AFQT score (Table 3.5).

[25]In these logistic regression models, $b = 0.043$ (0.015 standard error) for 240 groups of AIT graduates, whereas $b = 0.042$ (0.024 standard error) for 84 groups of unit members (see App. A, Tables A.10 and A.11). Although the effect of average AFQT score is statistically significant among AIT graduates, the coefficient for average AFQT score among unit members does not reach the conventional significance level ($p < .09$). We believe this result is due to the estimate having a larger standard error, which is associated with the much smaller sample size for unit members. Given the similarity of effects reflected in the AFQT score coefficients, we felt pooling was justified.

[26]Based on a log likelihood ratio test, chi-square = 4.41 with five d.f. (n.s.).

[27]Predictions from logistic regression models are more informative than simple empirical relationships, because they show how performance varies with the parameter of interest while controlling for values of other predictor variables. For example, when we examined the simple point-biserial correlation between average AFQT score and successful performance ($r = 0.15$, $p < .01$), the relationship appeared modest and uninformative as to differences in performance between the AFQT values of interest. As will be seen, substantial differences in performance were manifested when predictions were made for values of the parameter of interest while controlling for additional variables that influence the relationship between AFQT score and performance.

Unit Members

Figure 3.1 shows the probability that a three-person group of operators from an active-duty signal battalion will successfully operate the system for different average AFQT scores. The predictions were made for each group individually, with all predictors evaluated at the group's own value, except for aptitude, for which the midpoint of an AFQT category was used. The probability of success for each group was estimated using the logistic regression equation, and the average probability of success was computed across all groups. At each value of AFQT score examined, the result represents the expected probability of success for the groups in the sample with all factors except average AFQT score held constant.

In Fig. 3.1, the change in performance that would be expected from groups of different aptitude is the difference between the predictions at each AFQT category midpoint. For example, the results of this analysis suggest that for randomly selected groups of three unit members, if the average AFQT score is at the midpoint of category IIIA, 63 percent of these groups will successfully operate the system. If the average AFQT score of the groups is reduced to the midpoint of category IIIB, the probability of success will decline to 47 percent, a reduction of 16 percentage points.

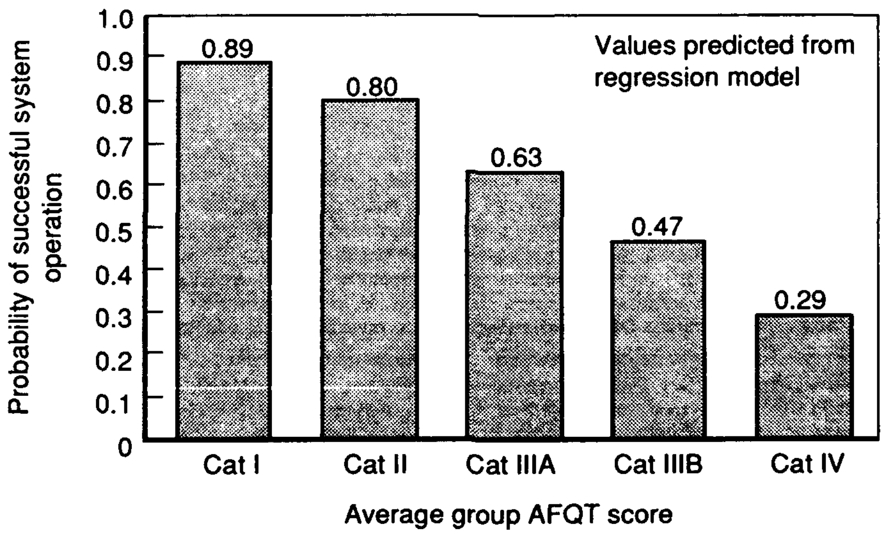

Fig. 3.1—System Operation and Group Aptitude, Unit Members

The predictions also indicate that if the average AFQT score of the group rises from the midpoint of category IIIA to the midpoint of category II, the expected probability of success will rise by 17 percentage points (from 63 to 80 percent). For larger changes in AFQT score, the performance differences will be correspondingly greater; for example, between the midpoints of categories II and IIIB, the difference in performance is estimated to be 33 percentage points. Thus, the model predicts a sharp increase in performance as the average AFQT score of the group increases across the range of the data.

Figure 3.1 also shows the estimated likelihood of success for groups whose mean AFQT scores are at the midpoints of categories I and IV. The differences appear very large, but caution should be exercised in interpreting results for such groups. These predictions are extrapolated from the model, with few groups in the sample actually found in these categories.

AIT Graduates

The predicted performance for operator groups composed of AIT graduates is shown in Fig. 3.2. Again, we see substantial differences in predicted performance across groups of varying AFQT levels, simi-

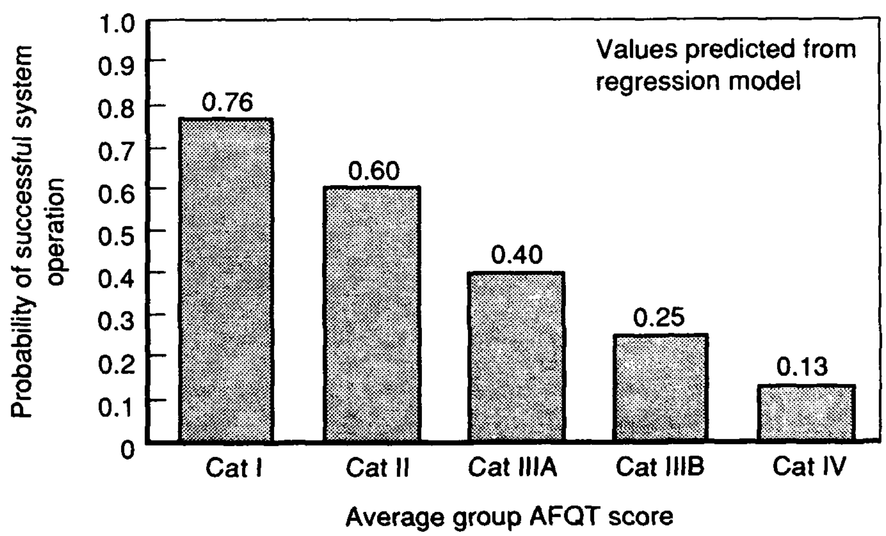

Fig. 3.2—System Operation and Group Aptitude, AIT Graduates

lar to the patterns among unit members. For example, among AIT graduates, the modal mean group AFQT score is within category IIIA. At that level, about 40 percent of groups are expected to succeed at system operation. If the average AFQT score of the group falls to the midpoint of category IIIB, however, we expect only 25 percent of such groups to successfully operate the system, which is a drop of 15 percentage points. Note that stated in relative terms, this difference is quite large—i.e., if the baseline rate is 40 percent and then drops by 15 points, the number of operating systems will be reduced by more than one-third. These results suggest that variations in group AFQT levels lead to very substantial differences in performance.

Predictions for High-Aptitude Members

Figures 3.3 and 3.4 show the predicted performance of groups of unit members and AIT graduates based on the model (shown in Table 3.7) that uses the number of high-aptitude members as the measure of group aptitude. As can be seen in each histogram, as the number of high-aptitude members increases, performance is expected to improve. The magnitude of the difference is similar in both models. Each high-scoring member added to the group increases the probability of group success by approximately 8 percentage points. For example, the model for unit members predicted that a group containing one high-aptitude member will succeed in 55 percent of the cases, whereas a group containing two high-aptitude members will succeed in 63 percent. The expected difference in performance is similar for AIT graduates (34 percent versus 42 percent, respectively).

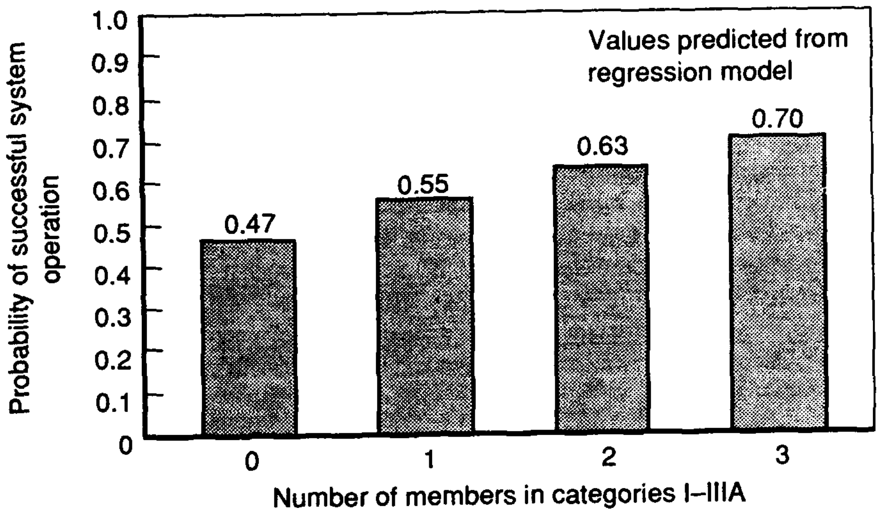

Fig. 3.3—System Operation and Group Composition, Unit Members

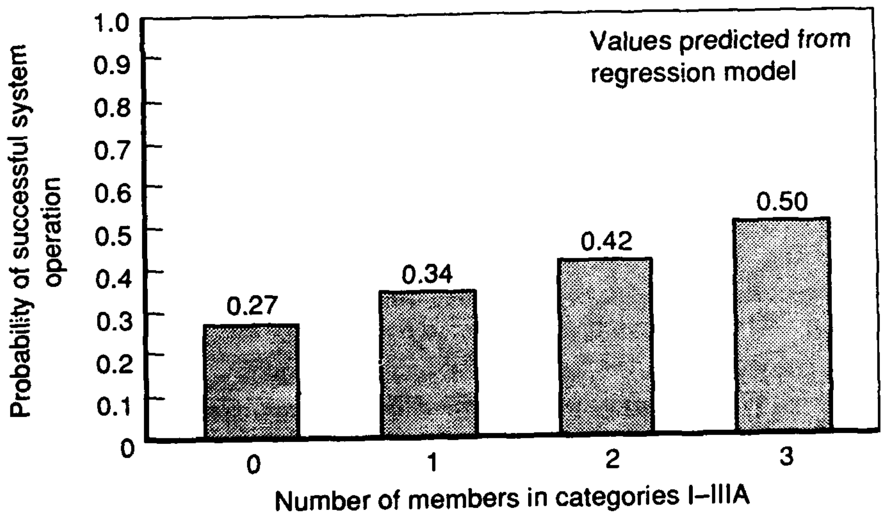

Fig. 3.4—System Operation and Group Composition, AIT Graduates

4. PREPARATION OF EQUIPMENT FOR OPERATION

Thus far, we have been primarily interested in the effects of member AFQT score on group performance in operating communications systems. However, questions are frequently raised about how AFQT score might affect individual proficiency on tasks required to operate communications equipment. It is difficult to assess individual proficiency in the group test because an operational communications system is the product of group interaction. Performance at one assemblage depends in large part on performance at the others. The design of our group test did not permit us to use the resulting data to examine how well each individual performed independent of the others. We were fortunate, however, in being able to draw upon earlier research at the Signal Center to examine the relationship between AFQT score and individual proficiency on communications tasks. This section presents results from our reanalysis of data collected in previous research.

The data were drawn from two studies carried out by RAND in 1986 through 1988 to examine the effectiveness of a novel training technology, interactive videodisc (IVD), for training communications operators in two occupational specialties.[1] The initial objectives of these studies were to examine the training effectiveness of alternative strategies for using IVD and similar training technology. The studies applied principles of controlled experimentation to compare effects of alternative methods used to train equivalent groups of soldiers. In both, the effects of traditional hands-on equipment training (the control condition) were compared with the effects of a training regimen using IVD (the experimental condition).

The studies involved members of two occupational specialties. The first study dealt with MOS 31M; the subsequent study dealt with MOS 31Q, Tactical Satellite/Microwave Systems Operator. Each study used approximately 340 students who were participating in Advanced Individual Training (AIT) at Fort Gordon. The students were, for the most part, entry-level personnel without prior military service. Their AFQT scores at entry, along with other demographic and individual characteristics, are shown in Table 4.1.

[1]For reports of these studies and detailed descriptions of their methodologies, see Winkler and Polich (1990).

Table 4.1

Characteristics of Trainees Participating in Earlier Experiments

Characteristic	31M Students	31Q Students
AFQT category (percent)		
I	3.6	2.4
II	31.6	43.6
IIIA	24.0	33.3
IIIB	36.7	20.4
IV	4.1	0.3
AFQT score, mean (s.d.)	56.8 (18.2)	62.2 (16.2)
Age (percent)		
17–19	54.1	45.3
20–21	22.2	23.9
22–23	13.6	13.6
24 or older	10.1	17.2
Age, mean (s.d.)	20.4 (3.1)	20.5 (2.1)
Sex (percent)		
Male	92.3	84.9
Female	7.7	15.1
Race (percent)		
White	66.9	75.5
Nonwhite	33.1	24.5
Education (percent)		
High school graduate	86.1	89.1
Nongraduate	13.9	2.6
Component (percent)		
Active	100.0	80.2
Guard or Reserve	0.0	19.8

NOTE: Based on 338 31M and 336 31Q AIT students.

Because these studies employed rigorous experimental designs and post-experimental performance assessments, and because we obtained the AFQT score of each study participant, they provided the basis for a subsequent analysis examining the effects of AFQT score on performance. Below, we describe the relationship between AFQT score and individual proficiency in each MOS, beginning with MOS 31M.

PRESET AND CABLING OF MULTICHANNEL TERMINALS

The first experiment evaluated the effects of IVD on student proficiency when used as a device to supplement hands-on equipment

training in MOS 31M. The experiment lasted seven months and covered 340 active-duty trainees who were assigned to one of two groups. In this study (as well as in the subsequent study, conducted in MOS 31Q), the experimental and control groups were formed using a statistical randomization model developed at RAND that provides a close match between groups on such factors as aptitude (AFQT score), educational background, demographic characteristics, and military experience.[2]

The amount and type of training received by each group were carefully monitored, and the performance of each trainee was subsequently assessed using the REES. This study focused on the initial steps of system operation, involving the preset and cabling of the most prevalent type of multichannel terminal, the AN/TRC-145. For this task, however, the REES was configured so that each node represented an independent terminal, and the performance of each terminal was individually evaluated by the REES computer as the operator concluded this specific task.

Using the computerized REES data, we analyzed individual performance using a measure analogous to that used in our group test: whether the trainee accomplished the preset and cabling task within the allowable time standard (15 minutes). Of the 337 soldiers with complete data, 183 (54 percent) completed the task correctly within the allowable time standard. This outcome was analyzed using regression models similar to those used in our analysis of group performance. That is, we examined the effects of AFQT score while controlling for differences in background (age, sex, race, education), amount and type of training (hands-on or hands-on supplemented with IVD), and REES test characteristics.[3] The means and standard deviations of variables used in this model are shown in Table 4.2.

The results of our regression analyses, shown in Table 4.3, indicate that AFQT score is statistically related to individual proficiency in presetting and cabling the AN/TRC-145.[4] That is, individuals with

[2]The method minimizes the variance of contrasts between groups on specific variables of interest, e.g., AFQT score. For discussion, see Press (1987) and Polich, Dertouzos, and Press (1986).

[3]These other measures included an indicator variable for the time of day in which testing was conducted (code 1 indicated testing in the afternoon or night shift) and the number of practice sessions on other assemblages in the REES facility provided to the trainee prior to testing.

[4]Presetting and cabling are the initial stages of system operation. The full operation task is the job that was performed by the team members in the test of group performance (see Sec. 3).

Table 4.2

Variables Used to Analyze Individual Performance of 31Ms

Variable	Mean	S.D.
Task completed to standard (yes/no)	0.54	0.50
AFQT score	56.79	18.27
Training indicator (IVD)	0.51	0.50
Sex (male)	0.92	0.27
Race (white)	0.67	0.47
Age	20.41	3.09
Education (high school graduate)	0.86	0.35
Shift of REES test (prime shift vs. off shift)	0.61	0.49
Practice time on REES before REES test (minutes)	65.90	79.81
Number of hands-on training sessions	9.41	2.80
Number of soldiers tested: 337		

Table 4.3

Regression Results for Terminal Preset Performance

Variable	Coefficient	Standard Error	Chi-Square
AFQT score	0.015	0.007	4.55*
Training indicator (IVD)	0.325	0.243	1.80
Sex (male)		0.452	0.44
Race (white)	0.045	0.269	0.03
Age	–0.055	0.040	1.87
Education (high school graduate)	–0.166	0.347	0.23
Shift of REES test (prime shift vs. off shift)	0.025	0.243	0.01
Practice time on REES before REES test	0.009	0.002	27.55*
Number of hands-on training sessions on task	0.002	0.044	0.00
Intercept	–0.444	1.225	0.13

NOTE: Logistic regression model used, based on 337 cases with complete data. *Model is significant* at $p < .001$ (chi-square = 42.88; $-2 \log L = 421.80$). Asterisk indicates parameter significant at $p < .05$.

higher AFQT scores are more likely than their lower-scoring counterparts to successful complete the presets and cabling of the assemblage within the time specified in the REES performance guide.

Predicted values from the logistic regression model estimating the probability of task completion are shown in Fig. 4.1. Again, we show mean values predicted by the model for a sample of operators scoring at the midpoint of each AFQT category. These estimates were obtained by evaluating the logistic function using individuals' values for all variables except AFQT score (the AFQT value at the midpoint of each category was substituted). Thus, the results represent the differences in performance that one would expect to observe based on the models for individuals in the sample, provided all factors except AFQT score are held constant.

As shown in the histogram, the model indicates that after controlling for other factors, the likelihood of completing the task to standard diminishes as AFQT score declines. For example, the results of this analysis suggest that in a randomly selected group of students, if all other factors are held constant, individuals scoring at the midpoint of category IIIB will be less likely to complete the task than individuals scoring at the midpoint of category IIIA (49 percent versus 54 percent, respectively), which is a difference of 5 percentage points. The predicted difference among groups scoring at the midpoints of categories II and IIIB is more dramatic—12 percentage points.

ALIGNMENT AND ADJUSTMENT OF RADIOS

Additional evidence on the relationship between AFQT score and individual proficiency was obtained in another RAND experiment, this one with AIT students in MOS 31Q, Tactical Satellite/Microwave Systems Operator.[5] That study examined the effects of varying the

[5]Members of MOS 31Q install, operate, and perform preventive maintenance checks and services and unit-level maintenance on tactical satellite, microwave/tropospheric scatter radios, associated multiplexing assemblages, and related equipment (antennas, generators, and communications security devices). Specialists are responsible for a variety of equipment involving different principles of propagation. Unlike the equipment used by MOS 31M, all of which employs line-of-sight, direct transmission of radio signals, the equipment used by MOS 31Q employs line-of-sight propagation (AN/TRC-138), tropospheric scatter (AN/TRC-112 or AN/TRC-121), or long-haul satellite links (AN/TSQ-85A and AN/TSQ-93A).

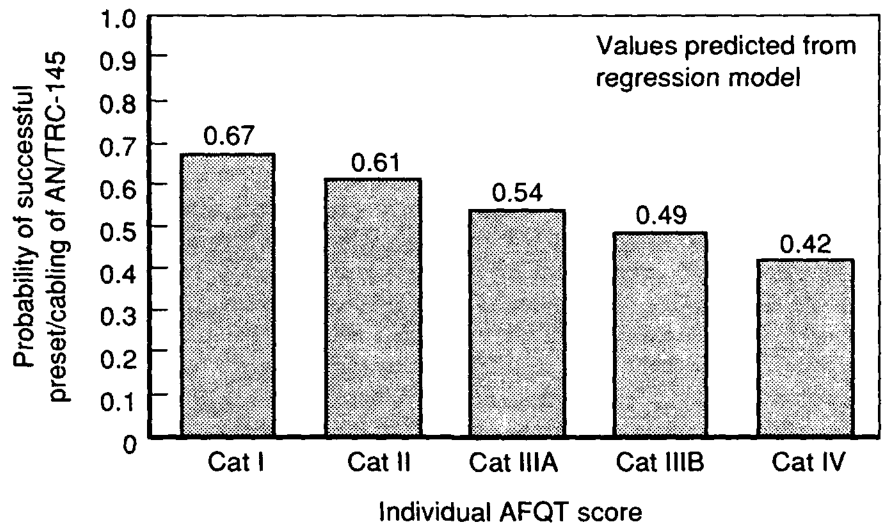

Fig. 4.1—Terminal Preset and Individual Aptitude, 31M AIT Students

mix of training resources: one group of students received training in a room containing a full complement of expensive tropospheric scatter radio equipment; the other group received equivalent training in a classroom where most (but not all) of the equipment had been replaced by IVD technology. Otherwise, the experimental method was similar to that of the 31M study, including balanced assignment of students to groups based on AFQT score and other individual characteristics, assessment of the amount and type of training received, and subsequent measurement of job-related performance.

This experiment, which lasted 11 months and involved 336 trainees, focused on training students in how to align and adjust tropospheric scatter (TROPO) radio assemblages.[6] It did not measure performance in the REES facility, however, because the REES facility does not contain the TROPO terminal set AN/TRC-121. Instead, the performance of each trainee was assessed via a hands-on test based on the 31Q soldier's manual for three relevant tasks (IF gain alignment, AGC alignment, and squelch adjustment). The hands-on tests were

[6]TROPO radios are generally aligned and adjusted each time the equipment in the communications shelter is prepared for operation, e.g., after movement in the field. These two tasks are required to prepare the equipment for subsequent operation in a communications system.

administered by objective assessors that had been trained and monitored by RAND, all of whom were unaware of the soldiers' AFQT scores or the methods used to train the soldiers. For each test, the assessors determined whether the trainee accomplished each of the tasks within the specified Army time standard.[7,8]

For the current study, we used logistic regression models to determine the relationship between individual AFQT score and the ability to complete each task according to the Army standard, again controlling for differences in individual background and type of training received.[9] The distributions of the variables used in the analyses are shown in Table 4.4. The results of the logistic regression analyses, shown in Table 4.5, indicate that AFQT score is a statistically significant predictor of performance on two of the tasks—AGC alignment and squelch adjustment—and a marginally significant predictor ($p < .07$) on the third—IF gain alignment.

To interpret these findings, we estimated the probability of successful task completion from the logistic regression equations in which AFQT score is a statistically significant predictor of performance.

Figure 4.2 displays probabilities estimated from the regression of AGC alignment for individuals at the midpoint of each AFQT category.[10] The model shows clearly that as aptitude decreases, performance is expected to decline. For example, the model estimates that the difference in the likelihood of correct alignment for individuals scoring at the midpoints of categories IIIB and IIIA will be 8 percentage points. Similar results are seen for the model of individual pro-

---

[7]The tests proved highly reliable for analytical purposes, with Cronbach's alpha, a measure of internal consistency, ranging from 0.80 to 0.95 over the three tasks. (See Winkler and Polich, 1990, for details.)

[8]Scheduling difficulties and the time required to conduct each test caused the number of students receiving each test to vary. These analyses are based on the following number of cases with complete data: IF gain (N = 323); AGC alignment (N = 296); squelch adjustment (N = 286). Tested populations did not differ in terms of aptitude or the other background variables used in the analysis.

[9]The predictor variables in these models were as follows: AFQT score, training method (code 1 = IVD/equipment mix; code 0 = equipment only), number of training sessions received on the task, age in years, and indicator variables for sex (1 = male), race (1 = white), and component (1 = active duty). We also included dummy variables to control for score variations attributable to the assessors.

[10]As in the logistic regression models presented previously, the function was evaluated at the individual's value for all predictor variables except AFQT score, for which the value at the AFQT category midpoint was used as the individual's value. Results were then averaged, and the probability of success was computed using the logistic function.

Table 4.4

Variables Used to Analyze Individual Performance of 31Qs

Variable	IF Gain		AGC Alignment		Squelch Adjustment	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Successful completion	0.34	0.47	0.41	0.49	0.83	0.38
AFQT score	62.16	16.17	62.16	15.87	62.99	15.69
Training indicator (IVD)	0.51	0.51	0.51	0.50	0.51	0.50
Number of training sessions	6.20	2.03	5.65	2.16	2.95	1.27
Age	20.51	2.08	20.56	2.08	20.61	2.08
Sex (male)	0.85	0.36	0.85	0.36	0.84	0.36
Race (white)	0.76	0.43	0.77	0.42	0.77	0.42
Component (regular Army)	0.82	0.38	0.82	0.38	0.81	0.39
Number of cases	323		296		286	

ficiency in squelch adjustment: as AFQT declines, so does the probability of completing the task to Army standard. Though students are more proficient on this latter task, the model estimates similar relationships between student AFQT scores and the likelihood of successfully completing the squelch adjustment to Army standard. Moreover, the size of the difference is consistent with that found for the model predicting performance at AGC alignment—i.e., 7 percentage points between individuals at the midpoint of category IIIA and individuals at the midpoint of category IIIB.

Table 4.5

Logistic Regression Results for Performance in TROPO Radio Alignments and Adjustments

Variable[b]	Ability to Complete IF Gain Alignment to Standard[a]			Ability to Complete AGC Alignment to Standard[b]			Ability to Complete Squelch Adjustment to Standard[c]		
	Coefficient	Standard Error	Chi-Square	Coefficient	Standard Error	Chi-Square	Coefficient	Standard Error	Chi-Square
AFQT score	0.015	0.008	3.49	0.025	0.009	8.57*	0.027	0.011	6.11*
Training indicator (IVD)	−0.034	0.243	0.02	0.063	0.260	0.06	−0.294	0.338	0.76
Sex (male)	0.575	0.371	2.41	0.352	0.375	0.88	0.407	0.454	0.80
Race (white)	0.257	0.300	0.73	−0.211	0.303	0.49	−0.058	0.394	0.02
Age	−0.056	0.060	0.89	−0.136	0.063	4.66*	−0.110	0.079	1.92
Component (active duty)	−0.101	0.312	0.10	0.482	0.341	1.99	0.694	0.395	3.08
Number of training sessions on task	.098	0.062	2.50	−0.108	0.064	2.84	0.122	0.139	0.77
Intercept	−1.703	1.438	1.40	0.950	1.454	0.43	0.406	1.889	0.05

NOTE: Asterisk indicates parameter significant at $p < .05$ (two-tailed test). Values for assessor indicator variables are not shown.

[a]Model significant at $p < .08$ (chi-square = 15.59, $-2 \log L$ = 398.76) based on 323 cases with complete data.

[b]Model significant at $p < .001$ (chi-square = 33.19, $-2 \log L$ = 367.25) based on 296 cases with complete data.

[c]Model significant at $p < .001$ (chi-square = 29.18, $-2 \log L$ = 235.92) based on 286 cases with complete data.

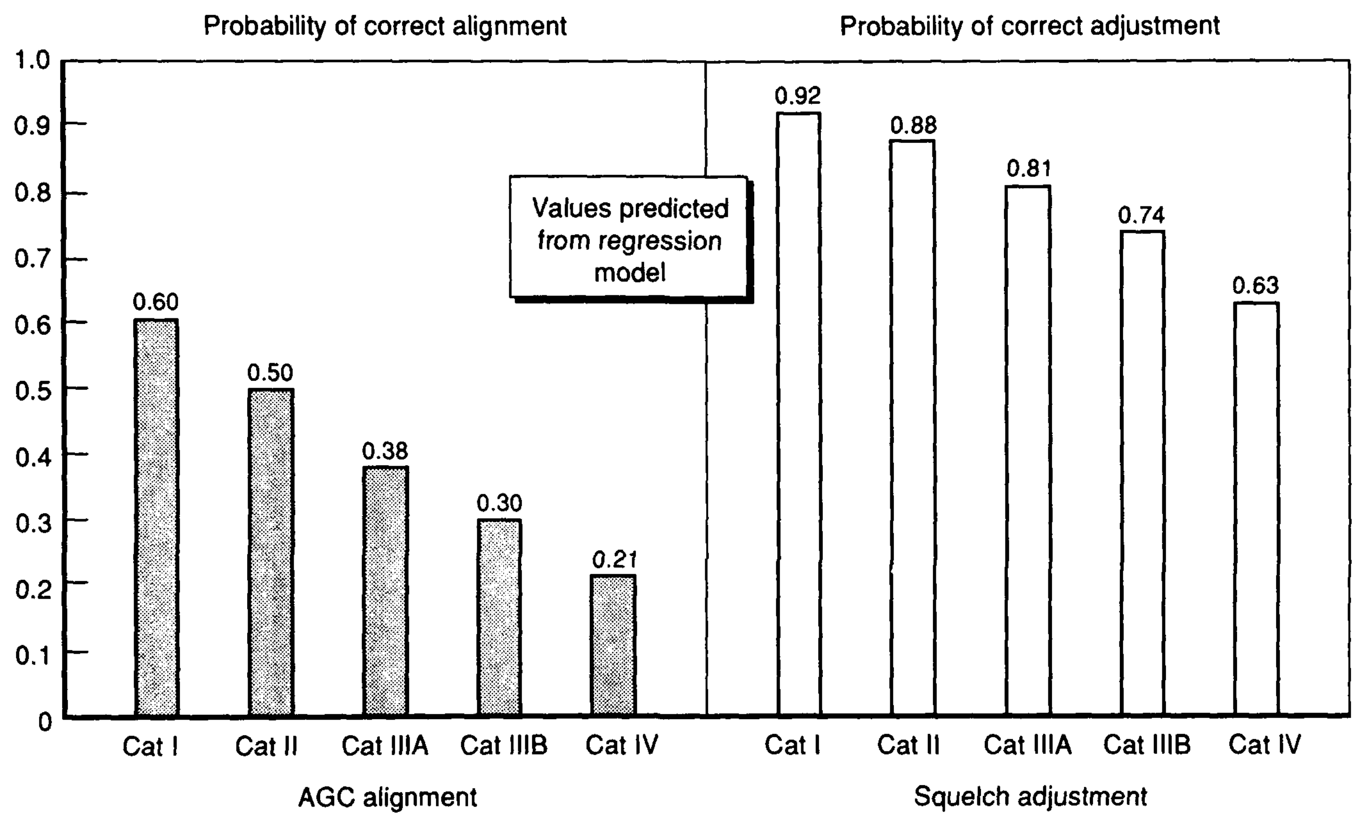

Fig. 4.2—AGC Alignment and Squelch Adjustment of TROPO Radio, 31Q AIT Students

5. SYSTEM TROUBLESHOOTING

In this section, we return to describing our evaluation of the performance of 31M soldiers interacting as a group in the REES facility, as described earlier, in Secs. 2 and 3. In this case, however, our analyses were aimed at assessing the link between aptitude and the ability of the group to troubleshoot an operating communications system that malfunctions and becomes disabled. We describe the test procedures first, followed by the estimated effects of AFQT score on group success in troubleshooting.

TESTING PROCEDURES

After completing the test of their ability to operate a communications network using the AN/TRC-145, soldiers were tested on their ability to troubleshoot malfunctions introduced into the communications system. The configuration was the same as in the operations test: two 12-channel systems connecting the person at the "control terminal" (terminal A) and the person at the "distant end" (terminal B), with a relay located in each system. Also, as in the system operation test, in one system (system 1) the relay was set to "override"; that is, the relay in that 12-channel system was unmanned, with the REES computer filling in for the absent relay operator. In the other system (system 2), the relay was operated by the third person in the test group.[1]

The REES can mimic the symptoms of a wide variety of malfunctions. In the field or in the REES, a soldier identifies a fault by performing a series of standard checks of his equipment, using a troubleshooting manual to determine the cause of any symptoms observed. In the field, the corrective action may consist of (a) repair or replacement of the faulty part by the operator or (b) referral of the equipment to higher-level maintenance. The nature and complexity of the fault determine which course of action is called for by the troubleshooting manual.[2]

[1]We distinguish between systems 1 and 2 in our discussion because the manned versus unmanned distinction determines where the symptoms of some faults appear.

[2]The official job description of the 31M soldier includes this "repair or replace" function for designated parts of the AN/TRC-145. Procedures established by communications units in the field, however, may discourage or forbid repair or replacement of any parts by the operators.

In the REES facility, no parts are repaired or removed. The soldier instead keys into the computer keyboard an alphanumeric code identifying the major piece of equipment with the malfunction, the malfunction's cause, and whether on-the-spot repair or higher-level maintenance is appropriate. The troubleshooting manual used in the REES is the same one used in the field, except for the inclusion of these alphanumeric codes.

The groups of AIT graduates tested in troubleshooting were the same ones tested in system operation, except when practical problems dictated that a team be used for only one part of the test.[3] A total of 239 trios of AIT graduates took the troubleshooting test. We also tested personnel serving in four active-duty signal battalions, but those data are not included here for two reasons: (1) the first-term soldiers in the units we tested apparently had performed only very limited troubleshooting, and (2) reliable data on the soldiers' troubleshooting experience were lacking.[4]

During the troubleshooting test, each group of soldiers underwent three separate trials. In each trial, the REES computer introduced the symptoms of two separate malfunctions into the communications network and the group was given 10 minutes to identify those malfunctions.

The specific malfunctions were selected by signal subject-matter experts, who recommended the use of multiple faults as a way to make the test more demanding for participants. (The malfunctions and their symptoms, together with the trial and station involved, are described in App. B.) No soldier's equipment had more than one malfunction in any trial, and in each trial the equipment of one group member had no malfunction. However, a soldier's equipment could exhibit a symptom due to a malfunction elsewhere in the communications network. This symptom could be in addition to symptoms produced by the soldier's "own" malfunction, or it could be from the malfunction of another soldier's equipment. The soldiers were not aware

[3]For example, because the REES facility is closed during thunderstorms, a group could complete its test in system operation and then have a thunderstorm prevent it from taking its troubleshooting test. Also, from time to time a computer malfunction resulted in the loss of data from a test group.

[4]The standard procedure in all four units from which test soldiers were drawn was that a first-term soldier calls for the help of an NCO for all but the simplest malfunctions. By unit policy, faults whose repair involves opening a piece of equipment were referred to unit maintenance personnel (a different MOS) for repair, regardless of whether the technical manual directs that the malfunction be repaired by an operator.

of the number or location of the malfunctions in any trial; all that was visible to them were the symptoms.

Once a fault was identified, the soldier with the faulty equipment had to diagnose the problem and key the relevant codes into the computer keyboard. A soldier might erroneously think a fault was present in his equipment (usually because a symptom of a fault elsewhere in the network manifested itself in his radio) and attempt to enter malfunction codes. These "false positives" did not count against the group; instead, we focused on the correct identification of faults inserted in the system.

Every malfunction used in the troubleshooting test had to be separately inserted into the REES computer by the console operator. Any malfunction not found by the soldiers within the 10-minute time limit for the trial had to be "cleared" by the console operator before the next trial. If it was not cleared, it would remain in the equipment, giving the next trial an unintended, extra malfunction. This process opened the possibility for console error, so we carefully examined the computer records for all cases in which console error could have occurred. Only trials in which both malfunctions were correct were included in the analysis. Console errors occurred in 6 percent of our trials, causing 18 percent of the AIT graduate groups tested to have less than three valid trials, as shown in in Table 5.1. We investigated possible bias imposed by the invalid trials and concluded that there were no such patterns.[5]

Table 5.1

Number of Valid System Troubleshooting Trials for AIT Graduates

Valid Trials	Number of Groups	Percentage of Groups
3	197	82.4
2	36	15.1
1	4	1.7
0	2	0.8

NOTE: Based on 239 groups of AIT graduates.

[5]The console operators did not know the AFQT scores of the soldiers being tested. Nonetheless, we checked for correlation between console error and soldier AFQT score and found it to be quite low ($r = -0.06$). We also used categorical tests to explore whether teams with one or more invalid trials did better or worse on their remaining trials than teams with no invalid trials. No evidence of bias was found.

Characteristics of Personnel Tested

We were able to obtain AFQT scores and demographic data on all three members of 187 of the 197 trios of AIT graduates that underwent three valid trials during the troubleshooting test. These groups formed the basis of the analysis. The characteristics of the individuals who composed these groups and the distribution of aptitude across the groups are shown, respectively, in Tables 5.2 and 5.3.[6] The distributions of variables measuring aptitude, background, and experience differ very little from those of the AIT graduates tested in system operation. Again, the distributions of group aptitude and the number of "high-scoring" members show substantial variability for AFQT categories II through IIIB. The range of values of average AFQT score for a three-person group of AIT graduates in our sample was 34.3 to 90.3, with most groups (40 percent of the sample) having average AFQT scores between 50 and 60.

Measures of Performance

In the troubleshooting test, there were originally two candidates for the outcome measure: (1) success in identifying the malfunction and (2) the time needed for successful identification. Early in the analysis, however, it became apparent that time-to-find was an unsatisfactory outcome measure. The relative paucity of cases in which the faults were successfully identified made the analysis difficult because of the small sample size. Further, comparison of the computerized records with data kept on site by project personnel suggested that other factors (e.g., typing skill) sometimes affected the speed and accuracy with which a soldier entered the lengthy malfunction codes.[7]

The malfunctions chosen for the test were relatively difficult to isolate, and task difficulty was further enhanced by the presence of more than one fault during each trial. These design principles were recommended by signal subject-matter experts as a way to ensure variability in the success rates among groups and to emphasize the importance of teamwork to group success. The test proved quite difficult

[6]Measures of team background and experience used in our analysis of troubleshooting performance were provided by the Defense Manpower Data Center, supplemented as necessary by data from Fort Gordon student personnel files.

[7]The system operation test merely required the soldier to press a "task-stop" button at the conclusion of the task. Thus, it is less likely that typing skill affected performance in that test.

Table 5.2

Characteristics of Individual 31Ms Tested in System Troubleshooting

Characteristic	Percentage of Total
AFQT category	
I	2.3
II	34.0
IIIA	24.2
IIIB	38.3
IV	1.1
Age	
18–19	15.9
20–21	41.0
22–23	14.0
24–25	15.9
26 or older	13.1
Sex	
Male	82.5
Female	17.5
Race	
White	67.1
Nonwhite	32.9
Education	
High school graduate	94.3
Nongraduate	5.7
Component	
Active	71.8
Guard or Reserve	28.3
Pay grade	
E1	76.8
E2	9.7
E3	11.8
E4 or greater	1.7

NOTE: Based on 561 AIT graduates who completed three valid troubleshooting trials and for whom AFQT data were available.

for examinees: groups customarily identified relatively few (two or less) of the six malfunctions introduced into their system, as shown in Table 5.4.

Table 5.3

Group Aptitude Distribution for System Troubleshooting Test

Group AFQT Composition		Percentage
Average AFQT score distribution by category and mean percentile		
I,	93–99	0.0
II,	65–92	24.1
IIIA,	50–64	55.1
IIIB,	31–49	20.9
IV,	10–30	0.0
Number of members in AFQT categories		
Three I–IIIA personnel		24.1
Two I–IIIA personnel, one IIIB–IV		39.6
Two or more IIIB–IV personnel		36.4

NOTE: Based on 187 three-person groups of AIT graduates who completed three valid troubleshooting trials and for whom AFQT scores were available.

Table 5.4

Group Performance in System Troubleshooting

Faults Correctly Identified	Percentage of Groups
0	12.8
1	26.7
2	29.9
3	22.5
4	8.0

Mean number of faults correctly identified: 1.9

NOTE: Based on 197 groups of AIT operators that had three valid troubleshooting trials. No group correctly identified more than four faults.

ANALYSES OF GROUP APTITUDE AND PERFORMANCE

Modeling Approach

As discussed in Sec. 3, there are several plausible hypotheses regarding how to represent group aptitude and how group aptitude affects performance. In preliminary investigations, we explored modeling approaches paralleling those used in our analysis of system operation,

using regression analysis with demographic variables and the following measures of aptitude:

- Average Ar'QT score of the group.
- Number of group members with AFQT scores in categories I through IIIA.
- Maximum or minimum AFQT score among the team members.
- AFQT score of each individual examinee, identified by the soldier's position within the communications net (terminal A, relay, or terminal B).
- Mean AFQT score of the two soldiers who had malfunctions inserted in their equipment in each trial.

Finally, as part of our preliminary regression analysis, we explored the possibility that AFQT score entered equations nonlinearly and examined other functional forms for representing group aptitude. The results were consistent with those found in our analysis of system operation. The average AFQT score of the group is a significant predictor of performance in troubleshooting. The results of the analyses suggest that the effects of individual AFQT scores on group outcome are additive, even though most of the contribution to group outcome on the troubleshooting task is made by the individual in the relay position.[8]

The final results presented here were based on an ordered polytomous logit analysis, using as the dependent variable the total number of malfunctions correctly identified by each group during three valid troubleshooting trials.[9] Polytomous logit with ordered outcomes has advantages over both ordinary least squares (OLS) analysis and other types of logit analysis.

[8]In addition to using AFQT score in various forms as a measure of aptitude, we explored the explanatory power of three other ASVAB scores—EL, SC, and GT—that are composite measures from the same battery of tests that produces AFQT scores. The EL score is designed to measure aptitude for electronics repair, the SC score to measure surveillance and communication aptitude, and the GT score to measure aptitude for general technical tasks. We placed special emphasis on EL scores because they are used to determine the eligibility of new recruits for MOS 31M. Regressions using EL (and SC and GT) had lower measures of goodness-of-fit than did regressions with AFQT alone; when both AFQT and EL were entered, AFQT was significant while EL was not.

[9]Dichotomous logit analysis is used when there are two possible outcomes, polytomous logit analysis when there are more than two outcomes, and ordered polytomous logit analysis when the multiple outcomes can be ranked from worst to best.

A scale of troubleshooting success could have been created by treating the number of malfunctions found as if it were a continuous variable and using simple OLS regression analysis. There would have been problems, however, in constructing such a variable and analyzing it using OLS regression. A continuous measure of troubleshooting success assumes that values of the outcome variable occur at equal intervals. For example, OLS regression would assume that finding four "bugs" is exactly twice as good as finding two. Our analyses showed, however, that some of the malfunctions used in the test were far easier to identify than others: the simplest fault was found by 64 percent of the groups tested, the hardest was found by only 19 percent (see App. A). In addition, teamwork might be much more important for identifying some malfunctions than it is for others. Thus, we felt that the assumption of equal intervals over malfunctions, implicit in a continuous measure, was less desirable than treating the dependent variable as an ordered measure with discrete values, in which case finding more bugs indicates "better" performance. Polytomous logit with ordered outcomes accomplishes this objective.

Analytic Results

The results of the ordered polytomous logit analysis clearly indicate that average group AFQT score has a statistically significant and positive effect on the troubleshooting performance of AIT graduates of MOS 31M. Several demographic variables were included in the regressions, as were two dummy variables. The first dummy variable indicated whether the manned relay in the communications net was in system 1 or system 2; the second indicated whether the AIT course was the one in effect before or after January 1, 1989, at which point the amount of troubleshooting training provided in the course was modified.[10]

All independent variables were entered linearly in the logit equations; the results of the logistic regressions are shown in Table 5.5.

The results of the logit analysis show average AFQT to be a statistically significant predictor of performance in troubleshooting. The average age of the team was also found to be significant: the sign on the

[10]The program of instruction in MOS 31M was modified to reduce the amount of troubleshooting training provided in the REES facility during the twelfth week of training. As this change in course syllabus did not affect system operation training, it was not included in that analysis.

Table 5.5

Regression of System Troubleshooting and Average Group AFQT Score

Variable	Coefficient	Standard Error	Chi-Square
Average group AFQT score	0.042	0.016	7.33*
Average age of operators	–0.134	0.069	3.81*
Number of high school graduates	0.502	0.315	2.54
Number of whites	–0.147	0.167	0.77
Number of males	0.151	0.200	0.57
Number of active-duty members	0.055	0.169	0.11
Test regimen (relay in system 2)	–0.931	0.350	7.07*
Course syllabus change	0.926	0.357	6.72*
Intercept 1	1.854	1.877	0.98
Intercept 2	0.238	1.877	0.02
Intercept 3	–1.101	1.875	0.34
Intercept 4	–2.768	1.885	2.16

NOTE: Ordered polytomous logit model used, based on 187 groups of AIT graduates with complete data. Model significant at $p < .02$ (chi-square = 18.64; $-2 \log L = 548.00$). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

coefficient indicates that younger teams perform better.[11] The number of high school graduates on a team was found to be positively related but not statistically significant ($p < .15$); sex, race, and component (active-duty versus Guard or Reserve) were not found to be significantly related to troubleshooting performance. Finally, both dummy variables were significant: isolating the particular faults we used was significantly more difficult with the manned relay in system 2, and the change in course content apparently improved team performance in troubleshooting.[12]

When individual AFQT scores of soldiers at each position are entered separately (in place of aggregate measures of group quality), the results of the polytomous logit analysis, shown in Table 5.6, indicate that the AFQT score of the soldier in the relay position is the major

[11]The use of a log formulation of age was also explored, as was estimation excluding groups at the extremes of the age distribution. Both approaches yielded results quite close to the ones reported in the test for linear, full-sample estimation.

[12]Given that the course syllabus reduced the amount of troubleshooting, this finding is surprising. However, we noted that the course received additional attention when the change occurred, and at about the same time, the Signal Center began to more closely monitor results emerging from the study. Thus, the "course effect," though it improves the precision of the model, should be interpreted with caution.

Table 5.6

Regression of System Troubleshooting and AFQT Score by Position

Variable	Likelihood of Successful System Operation		
	Coefficient	Standard Error	Chi-Square
AFQT score of terminal A operator	0.007	0.008	0.68
AFQT score of relay operator	0.028	0.009	10.35*
AFQT score of terminal B operator	0.008	0.008	1.03
Average age of operators	–0.130	0.069	3.52*
Number of high school graduates	0.517	0.315	2.69*
Number of whites	–0.134	0.168	0.64
Number of males	0.158	0.203	0.60
Number of active-duty members	0.103	0.172	0.35
Test regimen (relay in system 1)	–0.988	0.353	7.83*
Change in course syllabus	0.911	0.360	6.42*
Intercept 1	1.608	1.908	0.71
Intercept 2	–0.029	1.909	0.00
Intercept 3	–1.391	1.908	0.53
Intercept 4	–3.072	1.919	2.56

NOTE: Ordered polytomous logit model used, based on 187 groups of AIT graduates with complete data. Model significant at $p < .02$ (chi-square = 22.72; –2 log L = 543.92). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

contributor to group performance in troubleshooting. Thus, we cannot discount the alternative hypothesis (as we did in our analysis of system operation results) that group performance is dominated by the ability of one team member—in this case, the relay operator. We do not have enough information to judge whether this finding was an artifact of our selection of faults or a result of the fact that relay operations are less well practiced than are terminal operations in the 31M course, or whether AFQT score does indeed have a greater effect on troubleshooting success in the relay position. We do want to note, however, that all of the effects of individual AFQT score were positive and appear additive. Moreover, the performance evaluated was the group's performance, which from the commander's perspective is indistinguishable from the performance of any of the individuals in the group. Thus, if operators are randomly assigned to the three positions in this network, or if the assignment of soldiers to specific functions by their AFQT score cannot be carefully controlled in wartime, we think it would still be reasonable to predict group performance from an aggregate measure of group aptitude.

PREDICTED PERFORMANCE

Because the coefficients in ordered logit regression relate to the log of the odds of achieving success, the magnitude of the effect of a change in the value of an independent variable cannot be read directly from the estimated coefficients and will vary with the level of the variable. Therefore, instead of focusing further on the coefficient, we predicted the probability of success for groups at selected levels of aptitude. The process was the same one described in previous sections: for each team in the sample, all explanatory variables except the variable being evaluated took on their actual values. The evaluated variable was set at the same specified level for all teams. Then, the probability of success for each group was estimated using the logit equation, and the average probability of success across all groups was computed. The process was then repeated using a second value for the variable of interest. The difference in the two predicted probabilities was thus the difference in the odds of achieving success given all the actual characteristics of the soldiers in our model except for the variable of interest.

Table 5.7 shows the average probabilities predicted from the model that a group will correctly identify a minimum of one fault (or two, or three, or four faults) given the coefficients estimated by our logit regression. The rows report the predicted probability of identifying the

Table 5.7

Effect of Average Group AFQT on Troubleshooting Success of AIT Graduates

	Predicted Probability That Group Will Correctly Identify			
Average Group AFQT Score	1 or More Faults	2 or More Faults	3 or More Faults	4 Faults
Midpoint, Category I	97.1	97.4	66.0	28.9
Midpoint, Category II	94.3	77.5	49.4	16.7
Midpoint, Category IIIA	87.2	59.5	29.4	7.7
Midpoint, Category IIIB	77.5	42.9	17.4	4.0
Midpoint, Category IV	60.9	25.4	8.5	1.8
Actual value of average AFQT score (mean = 58)	87.4	60.3	30.5	8.2

indicated number of faults when a group is assigned all of its actual characteristics except for average group AFQT score, which was set at the score shown. The points chosen for evaluation here correspond to the midpoints of successive AFQT categories (I, II, IIIA, IIIB, and IV).

To illustrate our results, we focus on the predicted probabilities that groups with different average AFQT scores will correctly identify two or more faults—the median score among the groups examined. For AIT graduates, the chance of correctly identifying two or more faults (out of a possible six) is predicted to be about 60 percent for groups with an average AFQT score at the midpoint of category IIIA, which is roughly the mean AFQT score for soldiers in our sample. If the average AFQT falls to the midpoint of category IIIB, however, the chance of correctly identifying the same number of faults falls to 43 percent, a difference of 17 percentage points. The expected difference between groups falling at the midpoints of categories II and IV is even more extreme—35 percentage points.

These differences in troubleshooting performance are quite dramatic. We discuss their implications further in Sec. 6.

6. SUMMARY AND CONCLUSIONS

To assess the relationship between the aptitude and performance of signal operators, we analyzed the ability of groups and individuals to operate communications systems and the ability of groups to work together to isolate faults in a malfunctioning communications system. The results of our analyses show consistent and strong evidence that AFQT score is a significant predictor of job performance.

Estimates based on our multivariate models show that performance in system operation, an interactive task that determines the availability of communications in combat, is significantly related to the AFQT scores of the operators in the network. The results of our analyses show that after controlling for differences in group background and experience, average AFQT score is positively related to a group's ability to establish a multiple-node communications system. That is, groups with higher average AFQT scores are more likely to succeed at this task than are groups with lower average AFQT scores, and the latter are more likely to fail. In addition, the results indicate that the larger the *number* of high-scoring members in the group, the better the performance. These findings are true for both recent AIT graduates and members of active-duty signal battalions.

The results of related analyses of the individual proficiency of AIT graduates show consistent, statistically significant relationships between AFQT score and the performance of several duty-related tasks required to prepare communications assemblages for operation. Using data from earlier RAND research on AIT students at the Signal Center, we found that the higher a 31M operator's AFQT score, the more likely he will be to successfully accomplish the initial stages of system operation (preset and cabling of multichannel terminals). And the same is true for a 31Q operator of a tropospheric scatter radio: the higher his AFQT score, the more likely that he will prepare his equipment (i.e., perform alignments and adjustments) according to the technical standards prescribed by Army doctrine.

Performance in troubleshooting a communications system is also significantly related to the AFQT score of the operators in the network. The results of our analyses show that after controlling for differences in group background and experience, the average AFQT score of a three-person group of AIT graduates is positively related to the group's ability to isolate faults to the correct assemblage and component. Groups with higher average AFQT scores, and groups

with high-aptitude members in the relay position, find more of the "bugs" that interfere with communications.

Altogether, these results support the conclusion that "smarter" communications operators are better able to perform collective and individual tasks related to their wartime missions. In general, the effects are consistent and large. For example, our models suggest that for communications system operation, the effect of lowering the average AFQT score of a three-person group of unit members from the midpoint of AFQT category IIIA (near current levels) to the midpoint of category IIIB will be a decline of 16 percentage points in the probability that the command-and-control system will be successfully established. For AIT soldiers, the model estimates that the corresponding decline in performance will be 15 percentage points. These are large effects. Given baseline levels, such changes would reduce the number of correctly operated systems by more than one-third.

For system troubleshooting, our models suggest similarly large relationships between member aptitude and a group's ability to isolate malfunctions in communications systems. For example, our models suggest a decline of 17 percentage points in the probability of identifying two or more faults in a system (the median value) if the average AFQT score of a three-person group of AIT graduates falls from the midpoint of AFQT category IIIA (near current levels) to the midpoint of category IIIB.

We found the relationship between individual aptitude and proficiency to be smaller but consistent. The results of our analyses show that for the tasks and equipment examined, the lower the AFQT score of an operator, the greater the probability that assemblages will not be installed and aligned to standard. For example, the models suggest that if the AFQT score of operators falls from the midpoint of category IIIA to the midpoint of category IIIB, there will be a decline in performance ranging from 5 to 8 percentage points for each of the three tasks examined.

Based on the results of these analyses, we are confident that the relationship between the aptitude and job performance of enlisted personnel is consequential enough to affect the availability of battlefield communications. Our findings are relevant across a range of tasks and situations encountered by communications operators—for individuals preparing communications equipment for operation and for groups establishing and maintaining communications links among units. Moreover, our results for system operation show that group performance depends on the collective aptitude of the individual members and not on the ability of a single member. Because all

group members contribute and each additional higher-aptitude member improves group performance, changes in the overall aptitude mix could significantly affect the ability of communications operators to perform their collective missions.

The improvement in communications capability attributable to higher AFQT scores may be even more relevant during combat, when time pressure is acute and additional personnel, equipment, and maintenance support may well be in short supply. Thus, our results imply that should accession standards be lowered, we could expect a decrease in the successful performance of similar communications functions that provide command and control in wartime. The reduction in successful performance that these data imply will result from lowered AFQT standards should be considered in making budget and resource allocation decisions.

Appendix A

ALTERNATIVE MODELS OF APTITUDE AND PERFORMANCE IN OPERATING COMMUNICATIONS SYSTEMS

Table A.1

Regression of System Operation and Member Aptitude: Maximum AFQT Score in Group

Variable	Likelihood of Successful System Operation		
	Coefficient	Standard Error	Chi-Square
Maximum AFQT score in group	0.020	0.009	5.04*
Test population (unit members)	1.713	0.522	10.77*
Number of members using equipment	0.454	0279	2.65
Average age of operators	–0.118	0.058	4.18*
Number of high school graduates	0.064	0.250	0.06
Any Reservists in group	0.276	0.284	0.95
Number of males	0.043	0.175	0.06
Number of whites	0.128	0.150	0.73
Test regimen (relay in system 2)	–0.476	0.274	3.01
Intercept	–1.279	1.858	0.47

NOTE: Logit model used, based on 323 groups with complete data. Model significant at $p < .01$ (chi-square = 27.15; –2 log L = 418.37). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

Table A.2

Regression of System Operation and Member Aptitude: Minimum AFQT Score in Group

Variable	Likelihood of Successful System Operation		
	Coefficient	Standard Error	Chi-Square
Maximum AFQT score in group	0.026	0.011	5.24*
Test population (unit members)	1.778	0.527	11.37*
Number of members using equipment	0.473	0.279	2.89
Average age of operators	–0.078	0.057	1.89
Number of high school graduates	–0.006	0.249	0.00
Any Reservists in group	0.291	0.284	1.05
Number of males	0.093	0.178	0.27
Number of whites	0.168	0.148	1.29
Test regimen (relay in system 2)	–0.483	0.274	3.12
Intercept	–1.797	1.919	0.88

NOTE: Logit model used, based on 323 groups with complete data. Model significant at $p < .01$ (chi-square = 27.36; –2 log L = 418.16). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

Table A.3

Regression of System Operation and Member Aptitude: Average and Maximum AFQT Scores

	Likelihood of Successful System Operation		
Variable	Coefficient	Standard Error	Chi-Square
Average group AFQT score	0.046	0.020	5.32*
Maximum AFQT score in group	–0.004	0.014	0.10
Test population (unit members)	1.767	0.529	11.15*
Number of members using equipment	0.440	0.281	2.45
Average age of operators	–0.107	0.059	3.36
Number of high school graduates	0.026	0.253	0.01
Any Reservists in group	0.253	0.287	0.78
Number of males	0.140	0.181	0.60
Number of whites	0.087	0.153	0.32
Test regimen (relay in system 2)	–0.483	0.277	3.04
Intercept	–2.351	1.930	1.48

NOTE: Logit model used, based on 323 groups with complete data. Model significant at $p < .001$ (chi-square = 32.59; $-2 \log L = 412.93$). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

Table A.4

Regression of System Operation and Member Aptitude: Average and Minimum AFQT Scores

	Likelihood of Successful System Operation		
Variable	Coefficient	Standard Error	Chi-Square
Average AFQT score	0.042	0.019	5.02*
Minimum AFQT score in group	–0.000	0.016	0.00
Test population (unit members)	1.766	0.530	11.09*
Number of members using equipment	0.440	0.282	2.44
Average age of operators	–0.111	0.060	3.41
Number of high school graduates	0.034	0.253	0.02
Any Reservists in group	0.255	0.287	0.79
Number of males	0.134	0.181	0.55
Number of whites	0.083	0.153	0.30
Test regimen (relay in system 2)	–0.479	0.276	3.01
Intercept	–2.333	1.955	1.42

NOTE: Logit model used, based on 323 groups with complete data. Model significant at $p < .001$ (chi-square = 32.49; $-2 \log L = 413.02$). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

Table A.5

Regression of System Operation and Aptitude of Operator at Terminal A

	Likelihood of Successful System Operation		
Variable	Coefficient	Standard Error	Chi-Square
AFQT score of terminal A operator	0.146	0.007	4.81*
Test population (unit members)	1.765	0.529	11.15*
Number of members using equipment	0.483	0.281	2.96
Average age of operators	−0.101	0.056	3.23
Number of high school graduates	−0.010	0.249	0.00
Any Reservists in group	0.272	0.284	0.92
Number of males	0.034	0.174	0.04
Number of whites	0.150	0.149	1.02
Test regimen (relay in system 2)	−0.496	0.274	3.29
Intercept	−0.880	1.816	0.23

NOTE: Logit model used, based on 323 groups with complete data. Model significant at $p < .01$ (chi-square = 26.89; −2 log L = 418.62). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

Table A.6

Regression of System Operation and Aptitude of Operator at Relay Position

	Likelihood of Successful System Operation		
Variable	Coefficient	Standard Error	Chi-Square
AFQT score of relay operator	0.007	0.007	1.14
Test population (unit members)	1.656	0.518	10.23*
Number of members using equipment	0.482	0.277	3.05
Average age of operators	−0.096	0.056	2.93
Number of high school graduates	0.014	0.248	0.00
Any Reservists in group	0.270	0.283	0.91
Number of males	0.031	0.175	0.03
Number of whites	0.192	0.146	1.73
Test regimen (relay in system 2)	−0.511	0.273	3.52
Intercept	−0.686	1.836	0.14

NOTE: Logit model used, based on 323 groups with complete data. Model significant at $p < .01$ (chi-square = 23.14; −2 log L = 422.38). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

Table A.7

Regression of System Operation and Aptitude of Operator at Terminal B

Variable	Likelihood of Successful System Operation		
	Coefficient	Standard Error	Chi-Square
AFQT score of terminal B operator	0.012	0.007	3.38
Test population (unit members)	1.695	0.517	10.73*
Number of members using equipment	0.433	0.277	2.44
Average age of operators	–0.105	0.057	3.37
Number of high school graduates	0.056	0.249	0.05
Any Reservists in group	0.293	0.284	1.07
Number of males	0.041	0.175	0.06
Number of whites	0.180	0.147	1.50
Test regimen (relay in system 2)	–0.484	0.274	3.14
Intercept	–0.781	1.824	0.18

NOTE: Logit model used, based on 323 groups with complete data. Model significant at $p < .01$ (chi-square = 25.41; $-2 \log L = 420.10$). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

Table A.8

Regression of System Operation and Member Aptitude with Quadratic Term Included

Variable	Likelihood of Successful System Operation		
	Coefficient	Standard Error	Chi-Square
Average group AFQT score	–0.115	0.106	1.17
Average group AFQT score squared	0.001	0.001	2.17
Test population (unit members)	1.753	0.529	10.96*
Number of members using equipment	0.408	0.280	[illegible]
Average age of operators	–0.121	0.059	4.19*
Number of high school graduates	0.026	0.253	0.01
Any Reservists in group	0.287	0.290	0.98
Number of males	0.142	0.182	0.61
Number of whites	0.083	0.153	0.30
Test regimen (relay in system 2)	–0.441	0.279	2.50
Intercept	–2.331	3.662	0.41

NOTE: Logit model used, based on 323 groups with complete data. Model significant at $p < .001$ (chi-square = 34.76; $-2 \log L = 410.76$). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

Table A.9

Regression of System Operation and Member Aptitude with Interaction Terms Included

Variable	Likelihood of Successful System Operation: Coefficient	Standard Error	Chi-Square
Average group AFQT score	−0.277	0.215	1.17
Test population (unit members)	1.797	0.536	11.24*
Number of members using equipment	0.421	0.284	2.20
Average age of operators	−0.112	0.059	3.62
Number of high school graduates	0.008	0.255	0.00
Any Reservists in group	0.261	0.291	0.81
Number of males	0.146	0.182	0.65
Number of whites	0.087	0.154	0.32
Test regimen (relay in system 2)	−0.466	0.278	2.80
Terminal A/relay AFQT score interaction	0.002	0.001	1.65
Terminal B/relay AFQT score interaction	0.002	0.001	1.63
Terminal A/terminal B AFQT score interaction	0.002	0.001	1.76
Three-way AFQT score interaction	−0.000	0.000	1.15
Intercept	4.794	4.873	0.97

NOTE: Logit model used, based on 323 groups with complete data. Model significant at $p < .001$ (chi-square = 35.81; −2 log L = 409.70). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

Table A.10

Regression of System Operation and Group Aptitude: AIT Graduates Only

Variable	Likelihood of Successful System Operation: Coefficient	Standard Error	Chi-Square
Average AFQT score	0.043	0.015	7.57*
Average age of operators	−0.170	0.076	5.02*
Number of high school graduates	0.164	0.305	0.29
Number of males	0.105	0.193	0.30
Number of whites	0.135	0.176	0.59
Intercept	−0.352	1.957	0.03

NOTE: Logit model used, based on 240 groups with complete data. Model significant at $p < .05$ (chi-square = 14.31; −2 log L = 310.29). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

Table A.11

Regression of System Operation and Group Aptitude: Unit Members Only

	Likelihood of Successful System Operation		
Variable	Coefficient	Standard Error	Chi-Square
Average AFQT score	0.042	0.024	3.05
Average age of operators	–0.027	0.010	0.07
Number of high school graduates	–0.390	0.469	0.69
Number of males	0.328	0.470	0.49
Number of whites	–0.073	0.295	0.69
Intercept	–0.951	3.764	0.06

NOTE: Logit model used based on 84 groups. Model not statistically significant (chi-square = 4.84; –2 log L = 107.72).

Table A.12

Regression of System Operation and Member Aptitude with Test Population Interaction Terms Included

	Likelihood of Successful System Operation		
Variable	Coefficient	Standard Error	Chi-Square
Average group AFQT score	0.040	0.016	6.57*
Test population (unit members)	–1.318	4.481	0.19
Number of members using equipment	0.424	0.288	2.16
Average age of operators	–0.190	0.078	5.95*
Number of high school graduates	0.341	0.324	1.11
Any Reservists in group	0.312	0.293	1.13
Number of males	0.105	0.195	0.29
Number of whites	0.141	0.180	0.62
Test regimen (relay in system 2)	–0.545	0.282	3.73
Test pop./average AFQT score interaction	0.005	0.030	0.03
Test pop./number of high school grads interaction	–0.683	0.583	1.37
Test pop./average age interaction	0.176	0.129	1.87
Test pop./number of males interaction	0.309	0.526	0.34
Test pop./number of whites interaction	–0.085	0.359	0.36
Intercept	–1.488	2.221	0.45

NOTE: Logit model used, based on 323 groups with complete data. Model significant at $p < .001$ (chi-square = 36.90; –2 log L = 408.61). Asterisk indicates parameter significant at $p < .05$ (two-tailed test).

Appendix B

MALFUNCTIONS USED IN TESTING TROUBLESHOOTING

Table B.1

Malfunctions Used in Testing Troubleshooting

Malfunction	Percent of Time Found[a]	Trial	Node with Malfunction	Other Nodes with Symptoms	Problem	Description of Symptoms
RCVR 040 Sys 1	19	1	Relay	None	Defective order wire	Call indicator light does not go on and buzzer does not ring when distant end rings. RCVR meter indication is normal at order wire. Voice communication is possible. Function can be diagnosed with audio check of eqmt; also found if distant end tries to ring.
XMTR 008 Sys 1	27	1	Terminal A	None	Defective pressure differential monitor	Overheat indicator light is on (red); all meter readings are normal.
TD660 004 Sys 1	33	2	Terminal B	None	Defective buzzer	Audible alarm fails; must be diagnosed by audio check (visual check is not enough). Causes no red lights in any nodes; other nodes cannot help diagnose.
RCVR 030 Sys 2	64	2	Relay	Terminal A or B	Defective electronic switch (2A4)	HIGH SIG light is on in own node (red light); red light is on in terminal B's TD660 and KG 27. No communication is possible.[b]

Table B.1—continued

Malfunction	Percent of Time Found[a]	Trial	Node with Malfunction	Other Nodes with Symptoms	Problem	Description of Symptoms
XMTR 030 Sys 1	21	3	Terminal B	None	Defective meter switch (5TR1A1S1)	No meter reading regardless of switch position. Red light is on in TD660 during same trial due to RCVR 031.
RCVR 031 Sys 1	22	3	Terminal A	Relay	Poor line connections or damaged line or *antenna*	Appreciable meter indication with switch at REFL POWER; transmitter meter indication is also appreciable with switch at REFL POWER. LOW SIG light is on in relay; red light on terminal B's TD660 is on.[b]

[a]Based on 187 groups of AIT graduates with fully valid trials.

[b]Fault where the position of the manned relay affects the location of an "exported" symptom.

REFERENCES

Campbell, John P. (ed.), "Project A: The U.S. Army Selection and Classification Project," *Personnel Psychology*, Vol. 43(2) (special issue), Summer, 1990.

Department of the Army, *Combat Communications Within the Division*, Army Field Manual 11-50, Headquarters, March 1977a.

Department of the Army, *Wire and Multichannel Reference Data*, Field Manual 24-25, Headquarters, June 1977b.

Department of the Army, *Combat Communications Within the Corps*, Army Field Manual 11-92, Headquarters, November 1978.

Department of the Army, *Military Occupational Classification and Structure*, Regulation 611-201, Washington, D.C., 1989.

Department of Defense, *Appropriations Bill, 1988: Report of the Committee on Appropriations*, Report 100-000, Washington, D.C., October 1987.

Donaldson, Benjamin L., *Enlisted Force Quality Study*, U.S. Army Signal Center, 1985.

Gould, Inc., *Operator and Organizational Maintenance Manual for Reactive Electronics Equipment Simulator (REES) Device A8H10*, Simulation Systems Division, 1981.

Graham, Scott E., *Soldier Performance Research Project: Armor Field and SIMNET Test*, Research Report 1541, U.S. Army Research Institute, Fort Knox Field Unit, Fort Knox, KY, 1990a.

Graham, Scott E., *The Effects of Soldier Quality on Simulated Tank Gunnery Performance*, Research Report 1515, U.S. Army Research Institute, Fort Knox Field Unit, Fort Knox, KY, 1990b.

Greene, William H., *Econometric Analysis*, MacMillan Publishing Company, New York, 1990.

Hackman, R. J., and G. C. Morris, "Group Tasks, Group Interaction Process, and Group Performance Effectiveness," in H. H. Blumbert et al. (eds.), *Small Groups and Social Interaction*, John Wiley, New York, 1983.

Horne, David K., Maureen E. Reilly, and Aaron W. Schopper, *Predicting MLRS Performance from Army Selection Criteria*, draft

report, U.S. Army Research Institute, Alexandria, VA, January 1990.

Kahan, James P., et al., *Individual Characteristics and Unit Performance: A Review of Research and Methods*, R-3194-MIL, RAND, November 1984.

LeGree, Peter, and Michael Sanders, *Preliminary Statistics Quantifying the Quality of Signal Soldier NCO Course*, paper presented at the National Science Center Outreach in Technology and Education, Augusta, GA, April 1990.

Maddala, G. S., *Introduction to Econometrics*, Macmillan Publishing Company, New York, 1988, pp. 84–85.

O'Brien, G. E., and A. G. Owens, "Effects of Organizational Structure on Correlations Between Member Abilities and Group Productivity," *Journal of Applied Psychology*, Vol. 53(6), pp. 525–530.

Orvis, B. R., M. T. Childress, and J. M. Polich, *Effect of Personnel Quality on the Performance of PATRIOT Air Defense System Operators*, R-3901-A, RAND, 1991.

Polich, J. Michael, James N. Dertouzos, and S. James Press, *The Enlistment Bonus Experiment*, R-3353-FMP, RAND, April 1986.

Press, S. James, "The MISER Criterion for Imbalance in the Analysis of Covariance," *Journal of Statistical Planning and Inference*, Vol. 17, 1987, pp. 375–388.

Schopper, Aaron W., Richard W. Johnson, and Alexandra A. Burley, *Individual Differences in Target Location and Identification of Appropriate Target-Related Shell-Fuze Combinations Among Fire Support Specialists*, draft report, U.S. Army Research Institute, Fort Sill Field Unit, Fort Sill, OK, March 1990.

Scribner, B. L., et al., "Are Smart Tankers Better? AFQT and Military Productivity," *Armed Forces and Society*, Vol. 12(2), Winter 1986, pp. 193–206.

Silva, Jay M., *Usefulness of the Armed Services Vocational Aptitude Battery for Predicting Performance of TOW Gunners*, draft report, U.S. Army Research Institute, Alexandria, VA, March 1990.

Steiner, I. D., *Group Process and Productivity*, Academic Press, New York, 1972.

Tziner, Aharon E., "Effects of Team Composition on Ranked Team Effectiveness: The Blocked Fractional Factorial Design," *Small Group Behavior*, Vol. 19(3), August 1988, pp. 363–378.

Tziner, Aharon, and Dov Eden, "Effects of Crew Composition on Crew Performance: Does the Whole Equal the Sum of Its Parts?" *Journal of Applied Psychology*, Vol. 70 (1), 1985, pp. 85–93.

U.S. Army Signal Center, *Performance Guide: Reactive Electronic Equipment Simulator*, June 1983.

U.S. Total Army Personnel Command, *Force Management Book FY89*, Alexandria, VA, December 1989.

Wigdor, A., and G. Green, *Job Performance Measurement in the Military: Issues in Setting Military Enlistment Standards*, National Research Council, Washington, D.C., 1989.

Winkler, J. D., and J. M. Polich, *Effectiveness of Interactive Videodisc in Army Communications Training*, RAND, R-3848-FMP, January 1990.